

MINISTÉRIO DA SAÚDE  
FUNDAÇÃO OSWALDO CRUZ  
INSTITUTO OSWALDO CRUZ

Doutorado no Programa de Pós-Graduação em Biologia Computacional e Sistemas

O PAPEL BIOLÓGICO DAS ENZIMAS ISO-FUNCIONAIS NÃO HOMÓLOGAS EM  
LINHAGENS DE *Escherichia coli*

LEANDRO DE MATTOS PEREIRA

Rio de Janeiro  
Outubro de 2014

de Mattos Pereira, Leandro.

O PAPEL BIOLÓGICO DAS ENZIMAS ISO-FUNCIONAIS NÃO HOMÓLOGAS EM LINHAGENS de Escherichia coli / Leandro de Mattos Pereira. - Rio de janeiro, 2014.

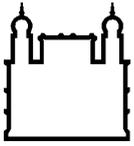
126 f.

Tese (Doutorado) - Instituto Oswaldo Cruz, Pós-Graduação em Biologia Computacional e Sistemas, 2014.

Orientador: Antonio Basilio de Miranda .

Bibliografia: f. 109-126

1. Enzimas Isofuncionais não homólogas Intragenômicas. 2. Nonhomologous gene displacement. 3. Escherichia coli. I. Título.



Ministério da Saúde

**FIOCRUZ**  
**Fundação Oswaldo Cruz**

## **INSTITUTO OSWALDO CRUZ**

**Programa de Pós-Graduação em Biologia Computacional e Sistemas**

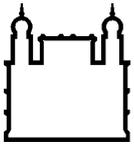
*LEANDRO DE MATTOS PEREIRA*

O papel biológico das enzimas iso-funcionais não homólogas em linhagens de *Escherichia coli*

Tese apresentada ao Instituto Oswaldo Cruz como parte dos requisitos para obtenção do título de Doutor em Biologia Computacional e Sistemas

**Orientador:** Prof. Dr. Antonio Basílio de Miranda

**RIO DE JANEIRO**  
**Outubro de 2014**



Ministério da Saúde

**FIOCRUZ**

**Fundação Oswaldo Cruz**

## **INSTITUTO OSWALDO CRUZ**

**Programa de Pós-Graduação em Biologia Computacional e Sistemas**

***AUTOR: LEANDRO DE MATTOS PEREIRA***

### **O PAPEL BIOLÓGICO DAS ENZIMAS ISO-FUNCIONAIS NÃO HOMÓLOGAS EM LINHAGENS DE *Escherichia coli***

**ORIENTADOR: Prof. Dr. Antonio Basílio de Miranda**

**Aprovada em: \_\_\_\_ / \_\_\_\_ / \_\_\_\_**

#### **EXAMINADORES:**

**Prof. Dra. Nome - Presidente Ana Carolina Paulo Vicente (FIOCRUZ)**

**Prof. Dr. Nome – Fábio Faria Mota (FIOCRUZ)**

**Prof. Dr. Nome - Ernesto Raúl Caffarena (FIOCRUZ)**

**Prof. Dr. Nome - Marcelo Alves Ferreira (FIOCRUZ)**

**Prof. Dr. Nome – Alberto Martín Ribeiro D'ávila (FIOCRUZ)**

Rio de Janeiro, 24 de Outubro de 2014

**Dedico aos meus amados pais, ao meu querido irmão e a todos os professores que me deram a oportunidade de me tornar uma pessoa e um profissional melhor.**

## AGRADECIMENTOS

Ao dom da vida, e a oportunidade que tenho todos os dias de aprender, à curiosidade e a busca insaciável por conhecimento que me aflinge.

Ao grande mestre Antônio, pela oportunidade, pelo aprendizado no âmbito profissional e pessoal, pela amizade, pelos conselhos, por toda experiência e aprendizado passado, por me ensinar ser um profissional melhor e pelo tratamento respeitoso.

Aos meus amados pais, Waldelírio e Leni. Pelo exemplo de vida e apoio. Agradeço de maneira especial a minha mãe por inconscientemente me inspirar na busca incansável pelo conhecimento, e por depositar em mim a educação que nunca teve.

Á meu irmão Leonardo Mattos pelos incentivos, conselhos, amizade e por sempre acreditar nos meus projetos e sonhos.

A minha querida namorada Ana Paula de Oliveira por todo apoio e por me amar em todos momentos.

A amiga Rangeline e amigos Marcos Catanho, Rafael Cuadrat, Márcio Albuquerque e Erico. Aos amigos e professores do Laboratório de Biologia Computacional e Sistemas.

Ao meu amigo Carlos Eduardo Monteiro, pela convivência, amizade, companheirismo e cumplicidade.

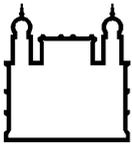
Aos professores, Ana Carolina, Fábio Mota, Ernesto Caffarena, por aceitarem o convite de fazer parte da banca examinadora, pelas correções, críticas e sugestões. Ao revisor da dissertação e suplente interno da banca Marcelo Alves pelas correções, sugestões, críticas e disponibilidade.

A todos os amigos e professores do Laboratório Alberto Dávilla, Fábio Mota, Rodrigo Jardim, Fabio Bernardo e todos os demais que contribuíram de maneira fundamental para minha formação profissional e realização deste trabalho.

A fonte financiadora Capes pela bolsa concedida durante o curso, que foi indispensável para o apoio logístico.

**“É impossível para o homem aprender aquilo que ele acha que já sabe”**

**Epiteto**



Ministério da Saúde

FIOCRUZ  
Fundação Oswaldo Cruz

## INSTITUTO OSWALDO CRUZ

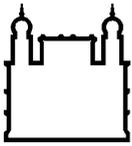
O papel biológico das enzimas iso-funcionais não homólogas em linhagens de  
*Escherichia coli*

### RESUMO

#### TESE DE DOUTORADO EM BIOLOGIA COMPUTACIONAL E SISTEMAS

Leandro de Mattos Pereira

Enzimas são proteínas que catalisam reações bioquímicas, funcionando como catalisadores biológicos naturais. Cada enzima tem um nome recomendado e número atribuído por uma comissão especializada, a "Enzyme Commission", um número que classifica a enzima de acordo com a reação que catalisa e de acordo com a sua função enzimática. As enzimas que catalisam a mesma reação geralmente têm uma alta semelhança entre as suas estruturas primária (por Ex. homólogos), isto reflete na semelhança das estruturas terciárias, o que sugere uma origem evolutiva comum. Contudo, algumas enzimas são capazes de catalizar a mesma transformação química e possuem distintas estruturas terciárias, são originadas de eventos evolutivos independentes, por processos de convergência funcional. Estas enzimas são conhecidas como enzimas isofuncionais não homólogas isofuncionais (NISEs). Analisando dados genômicos de diferentes linhagens de *Escherichia coli*, observamos a presença simultânea de NISEs no mesmo genoma. Não sabemos o significado biológico ou o impacto sobre o metabolismo da presença destas enzimas no mesmo genoma. Neste trabalho nós mostramos que estas enzimas têm diferentes estruturas terciárias confirmados pela validação estrutural e várias outras características funcionais diferentes, por exemplo, afinidade diferente para um substrato, diferentes padrões cinéticos e co-fatores, diferenças na estabilidade e regulação. Estas enzimas também podem ter diferentes padrões de expressão gênica e co-expressão e regulação, diferentes parceiros nas redes de interações proteína-proteína. A análise do número de cópias mostrou um aumento do número de cópias (Ex. amplificação gênica) de enzimas do metabolismo central envolvidas em processos essenciais, tais como: recombinação genética: (RusA: Crossover junction endodeoxyribonuclease RusA, EC 3.1.22.4), a modificação da parede celular e morte celular programada (RrrD: Lysozyme RrrD, EC 3.2.1.17), a subversão da resposta imune (UshA, 5'-nucleotidase, EC 3.1.3.5), formação de biofilmes, células de persistência (PlsB: Glycerol-3-phosphate acyltransferase, EC 2.3.1.15). Os resultados mostram que NISEs em linhagens de *E. coli* não representam casos de redundância funcional, tais enzimas aumentam a adaptação e flexibilidade metabólica da bactéria para lidar com diferentes desafios ambientais.



Ministério da Saúde

**FIOCRUZ**  
**Fundação Oswaldo Cruz**

## **INSTITUTO OSWALDO CRUZ**

The biological role of non-homologous iso-functional enzymes in strains of *Escherichia coli*

### **ABSTRACT**

#### **PHD THESIS IN BIOLOGIA COMPUTACIONAL E SISTEMAS**

**Leandro de Mattos Pereira**

Enzymes are proteins that catalyze biochemical reactions, functioning as natural biological catalysts. Each enzyme has a recommended name and number assigned by a specialized committee, the "Enzyme Commission", a number that ranks the enzyme according to the reaction that it catalyzes and according to its enzymatic function. The enzymes that catalyze the same reaction usually have a high similarity between their primary structures (by Ex. homologous) this similarity is reflected in tertiary structures, suggesting a common evolutionary origin. However, some enzymes are able to catalyze the same chemical transformational and have distinct tertiary structure, are originated from distinct events evolutives by process of functional convergence. Such enzymes are known as non-homologous enzymes isofuncionais (NISEs) or similar enzymes. Analyzing genomic data of different strains of *Escherichia coli*, we observed the simultaneous presence of NISEs the same genome. Do not know the biological significance or the impact on the metabolism of the presence of these enzymes in the same genome. In this work we have shown that these enzymes have different tertiary structures confirmed by structural validation, and have several other different functional features, for example: different affinity for a substrate, different kinetic patterns and co-factors, different stability and regulation. We also observed that these enzymes may have different expression patterns and gene co-expression and regulation, different partners in the networks of protein-protein interactions. The analysis of the number of copies showed an increased number (Ex. gene amplification) of copies of some structural forms with analogy cases of central metabolism, enzymes involved in essential processes such as: genetic recombination: (RusA: Crossover junction endodeoxiribonuclease, EC 3.1.22.4), modification of cell wall and programmed cell death (RrrD: lysozyme), subversion of the immune response (Usha, 5'-nucleotidase, EC 3.1.3.5) formation biofilms and cell persistence (PlsB: acetyl glicerol 3-phosphatase, EC 2.3.1.15). The results show that NISES don't represent cases of functional redundancy; such enzymes increase metabolic flexibility of bacteria to handle different environmental challenges.

## SUMÁRIO

<b>1 INTRODUÇÃO.....</b>	<b>15</b>
1.1 As enzimas isofuncionais não-homólogas.....	15
1.2 Enzimas e sua classificação.....	19
1.3 <i>Escherichia coli</i> e pangenoma.....	21
1.4 Biologia de sistemas.....	23
<b>2 OBJETIVO GERAL.....</b>	<b>27</b>
2.1 Objetivos Específicos.....	27
<b>3 METODOLOGIA.....</b>	<b>28</b>
3.1 Importação de dados do KEGG e predição computacional das formas análogas.....	28
3.2 Identificação de análogos intragenômicos em <i>E. coli</i> .....	29
3.3 Classificação dos enovelamentos estruturais e superfamílias das proteínas preditas presentes nos proteomas das linhagens de <i>E. coli</i> .....	30
3.4 Modelagem por homologia e análise estrutural comparativa.....	31
3.5 Identificação e análise de cofatores, substratos, parâmetros cinéticos, inibidores, localização celular, parâmetros físicos químicos e bioquímicos.....	32
3.6 Análises de rede de interação dos análogos intragenômicos em <i>E. coli</i> .....	33
3.7 Análises de expressão e co-expressão gênica.....	34
3.8 Análise do contexto genômico.....	34
3.9 Análise dos fenótipos de mutantes em <i>Escherichia coli</i> , essencialidade, resistência para antibióticos e <i>fitness</i> .....	35
<b>4 RESULTADOS.....</b>	<b>36</b>
4.1 Identificação de análogos funcionais em <i>E. coli</i> .....	38

4.2 Superóxido dismutase (EC 1.15.1.1): SodA, SodB, SodC.....	47
4.2.1 Número de cópias e contexto genômico.....	47
4.2.2- Análise do perfil de expressão gênica de sodA, sodB e sodC.....	50
4.2.3 Análise dos dados e literatura.....	51
4.3 Dihidrofolato redutase (EC 1.5.1.3): FoIA e FoIM.....	53
4.3.1 Análise da taxa de sobrevivência de mutantes ( <i>E. coli</i> ) <i>foIA</i> e <i>foIM</i> quando submetidas à diferentes estresses ou compostos químicos.....	56
4.3.2 Parâmetros físico-químicos e comparação estrutural entre FoIA e FoIM.....	58
4.3.3 Análise dos dados e literatura.....	61
4.4 <i>Crossover junction endodeoxiribonuclease</i> (EC 3.1.22.4): <i>rusA</i> , <i>ruvC</i> .....	65
4.4.1 Análise da taxa de sobrevivência de mutantes ( <i>E. coli</i> ) <i>rusA</i> ou <i>ruvC</i> quando submetidas à diferentes estresses ou compostos químicos.....	67
4.4.2 Comparação estrutural entre RusA e RuvC através de modelagem comparativa e identificação dos resíduos do sítio ativo e sítio alostérico.....	69
4.4.3 Análise dos dados e literatura.....	70
4.5 Glicerol-3-fosfato-1-O-aciltransferase (EC 2.3.1.15): <i>plsX</i> , <i>plsB</i> e <i>plsY</i> .....	76
4.5.1 Rede de interação, contexto genômico e essencialidade dos genes <i>plsX</i> , <i>plsB</i> e <i>plsY</i> .....	77
4.5.2 Análise da taxa de sobrevivência de mutantes <i>plsB</i> , <i>plsX</i> e <i>plsY</i> quando submetidas à diferentes estresses ou compostos químicos.....	81
4.5.3 Análise dos dados e literatura.....	82
4.6 5'-Nucleotidase (EC 3.1.3.5): <i>SurE</i> , <i>UshA</i> , <i>YfbR</i> , <i>YjjG</i> .....	88
4.6.1 Número de cópias de <i>surE</i> , <i>ushA</i> , <i>yfbR</i> e <i>yjjG</i> .....	88
4.6.2 Comparação estrutural e modelagem comparativa.....	91
4.7 Lisozima (EC 3.2.1.17): <i>rrrD</i> , <i>yefX</i> , <i>chiA</i> , endolisina.....	96
4.8 Arginina decarboxilase (EC 4.1.1.19): <i>SpeA</i> e <i>AdiA</i> .....	100

<b>5 DISCUSSÃO.....</b>	<b>102</b>
<b>6 CONCLUSÕES.....</b>	<b>108</b>
<b>7 REFERÊNCIA.....</b>	<b>109</b>

## LISTA DE FIGURAS

Figura 1 – Representação gráfica de algumas NISEs no genoma de <i>E. coli</i> K12.....	42
Figura 2 – Número de cópias de <i>sodA</i> , <i>sodB</i> e <i>sodC</i> nas 33 linhagens de <i>E. coli</i> estudadas.....	47
Figura 3 – Contexto genômico local de <i>sodC</i> nas linhagens K12, O157: H7 EDL933 (EHEC), O157: H7 Sakai (EHEC) e SMS-3-5 (ambiental).....	49
Figura 4 – Padrão de expressão gênica de <i>sodA</i> e <i>sodC</i> na presença e ausência de oxigênio.....	50
Figura 5 – Padrão de expressão gênica de <i>sodB</i> e <i>sodC</i> na presença e ausência de oxigênio.....	51
Figura 6 – Número de cópias de <i>folA</i> e <i>folM</i> nas 33 linhagens de <i>E. coli</i> estudadas. EHEC: Enterohemorrágica, EPEC: Enteropatogênica, ExPEC: Estraintestinal, ETEC: Enterotoxigênica.....	54
Figura 7 - Contexto genômico das duas cópias de <i>folA</i> em <i>E. coli</i> O104: H4 2009 EL-2050.....	55
Figura 8 – Contexto genômico de <i>folM</i> em algumas linhagens patogênicas e não patogênicas de <i>E. coli</i> . A descrição completa dos genes encontra-se na <i>FiguraS2</i> .....	55
Figura 9 – Taxa de <i>fitness</i> do mutante <i>folM</i> de <i>E. coli</i> BW25113 tratados com inibidores da síntese de tetraidrofolato.....	56
Figura 10 – Rede de interação de <i>FolA</i> obtida do banco de dados STRING.....	57
Figura 11 – Rede de interação de <i>FolM</i> obtida do banco de dados STRING.....	58
Figura 12 – Estrutura 3D de <i>FolA</i> e <i>FolM</i> .....	60
Figura 13 - Número de cópias de <i>rusA</i> e <i>ruvC</i> nas 33 linhagens de <i>E. coli</i> estudadas.....	65
Figura 14 - Contexto genômico das diferentes cópias do gene <i>rusA</i> em diferentes linhagens de <i>E. coli</i> .....	66
Figura 15 - Representação esquemática da via do metabolismo de glicerofosfolídeos obtida do KEGG com a ferramenta KEGG Mapper.....	77
Figura 16 - Rede de interação de <i>PlsB</i> . ....	78

Figura 17 - A: Contexto genômico de (a) <i>plsB</i> , (b) <i>plsX</i> e (c) <i>plsY</i> em <i>E. coli</i> K12.....	79
Figura 18 - Rede de interação de PlsX.....	80
Figura 19 - Rede de interação de PlsY.....	80
Figura 20 - A: Taxas de <i>fitness</i> dos mutantes <i>plsX</i> , <i>plsY</i> e <i>folA</i> para <i>Escherichia coli</i> BW25113 tratado com inibidores da síntese de proteínas e indutor de estresse Paraquat.....	81
Figura 21 - Número de cópias das formas análogas das 5'-nucleotidasas: SurE, UshA, YfbR e YjjG nas 33 linhagens estudadas.....	89
Figura 22 - Modelo 3D de SurE, UshA, YfbR e YjjG.....	93
Figura 23 - Número de cópias de <i>rrrD</i> (lisozima_1), <i>yefX</i> (lisozima_2), <i>chiA</i> (lisozima_3) e endolisina (lisozima_4) nas 33 linhagens estudadas.....	97
Figura 24 - Contexto genômico de lisozima_1 ( <i>rrrD</i> ) em <i>E. coli</i> O157:H7 Sakai (EHEC).....	99

## LISTA DE TABELAS

Quadro 1 - Classes enzimáticas e classificação segundo a União Internacional de Bioquímica.....	20
Tabela 1 - Linhagens de <i>Escherichia coli</i> cujos genomas foram utilizados neste trabalho.....	37
Tabela 2- Enzimas análogas intragenômicas em <i>E. coli</i> .....	40
Tabela 3 - Quadro comparativo dos métodos analíticos empregados neste trabalho.....	44
Tabela 4: Parâmetros bioquímicos das enzimas análogas <i>adiA</i> e <i>speA</i> .....	101

# 1 INTRODUÇÃO

## 1.1 As enzimas isofuncionais não-homólogas

Enzimas isofuncionais não homólogas (NISEs), ou enzimas análogas, são enzimas que catalisam a mesma reação enzimática, mas que possuem origens evolutivas independentes, e por um processo de evolução convergente estas enzimas desenvolveram a capacidade de catalisar a mesma transformação química (reação enzimática). Esta teoria é apoiada pelo fato destas enzimas não possuírem similaridade significativa em suas estruturas primárias e possuírem distintas estruturas terciárias. Os processos de evolução convergente em enzimas foram primeiramente descritos há 40 anos por Wright e colaboradores (Wright *et al.*, 1969) em seus estudos sobre a estrutura tridimensional de subtilisina (serina endopeptidase). Outros casos de enzimas análogas foram relatados, como por exemplo, as superóxido dismutases Cu/Zn- e Mn/Fe-dependente (Stallings, Powers, Pattridge, Fee, & Ludwig, 1983) e as  $\beta$ -lactamases do tipo I (classes A e C) e tipo II (Carfi *et al.*, 1995).

O primeiro estudo em larga escala abrangendo a identificação de enzimas análogas em diversos genomas foi realizado em 1998 por Galperin e colaboradores (1998) (Galperin, Walker, & Koonin, 1998), utilizando dados do GenBank ([www.ncbi.nlm.nih.gov/genbank](http://www.ncbi.nlm.nih.gov/genbank)). Neste estudo foi observado que algumas enzimas, após terem suas estruturas primárias comparadas com o programa Blastp (Stephen F Altschul *et al.*, 1997), obtiveram pontuação (*score*) abaixo de 120. Isto indica que nenhuma similaridade significativa entre suas estruturas primárias foi detectada. Por outro lado, estas mesmas enzimas comprovadamente compartilhavam a mesma atividade enzimática e apresentavam estruturas tridimensionais significativamente distintas, o que foi medido através de alinhamentos estruturais.

A partir desta evidência experimental, tornou-se possível a predição computacional de casos de analogia através da simples comparação das estruturas primárias de enzimas que compartilhem a mesma atividade enzimática, com o uso do programa de alinhamento local mais difundido no meio científico.

Na maioria das vezes a ausência de similaridade detectável entre as sequências primárias de proteínas é atribuída a uma grande divergência entre sequências homólogas que acompanha a adaptação para diferentes condições ambientais (Galperin & Koonin, 2012a), mas ao nível de comparação estrutural tridimensional ainda podemos observar sobreposição de suas estruturas terciárias quando alinhadas, compartilhando os mesmos enovelamentos, significando que estas, na maioria das vezes são homólogas

Em outros casos, algumas enzimas que catalisam a mesma transformação química possuem estruturas tridimensionais distintas e são classificadas em duas ou mais distintas superfamílias, o que indica que também devem ter surgido e evoluído de formas independentes.

Em 2008, Otto e colaboradores (Otto *et al.*, 2008) desenvolveram o *programa* AnEnPi que, com base nas evidências apresentadas no trabalho de Galperin e colaboradores (Galperin *et al.*, 1998), permite a predição computacional de casos de analogia a partir do conjunto de estruturas primárias de enzimas com atividade enzimática anotada (oriundas de genomas completamente sequenciados) disponíveis no banco de dados KEGG (<http://www.genome.jp/kegg/kegg1.html>). Neste trabalho os autores apresentaram 35 casos potenciais de analogia intergenômica (definida como a existência de analogia entre enzimas presentes em distintos genomas) entre o ser humano (*Homo sapiens*) e um dos parasitos causadores da leishmaniose (*Leishmania major*), assim como 12 casos possíveis de analogia intragenômica (definida como a existência de analogia entre enzimas presentes no mesmo genoma) neste parasito.

Em 2010, Omelchenko e colaboradores (Omelchenko, *et al.*, 2010) demonstraram que 185 atividades enzimáticas (ECs) distribuídas nos genomas sequenciados possuem duas ou mais proteínas com estruturas tridimensionais distintas capazes de catalisare a mesma transformação química, e no mesmo trabalho propuseram uma nova nomenclatura para estas enzimas, passando então a serem denominadas enzimas isofuncionais não homólogas (*Non-homologous Isofunctional Enzymes* - NISE). Em nosso trabalho usaremos os termos análogos funcionais como sinônimo de enzimas isofuncionais não homólogas.

Em 2010, Capriles e colaboradores (Capriles *et al.*, 2010) identificaram enzimas análogas intergenômicas entre *Homo sapiens* e o patógeno *Trypanosoma cruzi*. Considerando as diferenças estruturais encontradas entre as enzimas análogas, estas

possuem potencial para ser utilizadas como alvos para o desenvolvimento de novos fármacos. Isto significa que as enzimas análogas correspondem a um reservatório, até o momento pouco explorado, de potenciais alvos terapêuticos. De fato, alguns alvos terapêuticos apresentam casos de analogia, como, por exemplo, as enzimas tripanotiona redutase de *Leishmania donovani* e glutathiona redutase de mamíferos (Cunningham & Fairlamb, 1995).

Do ponto de vista biológico, a presença de duas enzimas com o mesmo papel funcional representaria uma duplicação ou redundância funcional, o que vai contra o paradigma de economia da energia metabólica, pois para cada proteína sintetizada existe um custo energético associado não somente à sua própria síntese, como também à síntese de seus aminoácidos. Em *Escherichia coli*, já foi observado que proteínas altamente expressas são deficientes em aminoácidos mais custosos do ponto de vista de sua síntese (Bragg & Wagner, 2007). Em um cenário evolutivo alternativo, as enzimas análogas funcionais podem ter adquirido funções adicionais, como funções estruturais. O fenômeno *moonlighting* é descrito para enzimas que adquirem funções adicionais não enzimáticas (Galperin & Koonin, 2012b).

Em 2011, Foster e colaboradores (Foster *et al.*, 2010) analisaram em *Escherichia coli* o papel funcional das NISES fosfoglicerato mutase dependente de cofator (dPGM) e independente de cofator (iPGM). As enzimas fosfoglicerato mutase catalisam a interconversão de 2- e 3- fosfoglicerato (2-PG e 3-PG) na via glicolítica e gluconeogênica. Neste estudo, foi observado que a atividade catalítica da enzima dPGM é maior que da sua forma alternativa iPGM. Em *Bacillus subtilis*, a deleção do gene iPGM resultou em fenótipo de crescimento e esporulação; deleções no gene dPGM não produziram nenhum efeito (Pearson *et al.*, 2000).

A presença de NISES em um mesmo genoma pode representar soluções alternativas para uma determinada condição biológica ou mesmo uma rota de escape em uma via metabólica, pois uma execução alternativa da mesma transformação química pode ocorrer proporcionando ao organismo possuidor destas formas enzimáticas alternativas uma maior flexibilidade metabólica. Portanto, torna-se plausível presumir que as mesmas possam ter grande importância bioquímica, desempenhando papéis importantes ou chaves em determinadas situações

Neste estudo, nós utilizamos como definição de trabalho para o termo “papel biológico” de uma determinada enzima como sua função global em um determinado contexto ou condição biológica. Em outras palavras, o “papel biológico” desta enzima compreende não somente sua atividade catalítica particular como também quaisquer funções estruturais (ou outras) que possa ter.

Apesar do fato de enzimas análogas catalisarem a mesma reação enzimática do ponto de vista bioquímico (número de EC), pretendemos mostrar que o papel biológico destas enzimas em um organismo pode, em algumas ocasiões, ser distinto, pois estas enzimas possuem estruturas completamente diferentes, assim como diversos outros parâmetros (cinética, cofatores, custo de síntese, etc.). Pretendemos mostrar que as enzimas análogas são alternativas para enfrentar desafios impostos pelas condições e mudanças ambientais.

Até o momento, a maioria dos trabalhos na literatura envolvendo a predição e a caracterização funcional e estrutural de proteínas tem negligenciado o estudo dos casos de NISES presentes em um mesmo genoma. A maioria dos trabalhos publicados tem como objetivo a procura de proteínas homólogas entre diferentes organismos. Neste trabalho, procuramos oferecer outro ponto de vista, contextualizando as diferenças entre a função enzimática e o papel biológico, utilizando estas enzimas como modelo de estudo.

Para um maior entendimento das questões supracitadas, nós estudamos o papel funcional das enzimas análogas intragenômicas através da análise de diversos dados, como por exemplo: i) o padrão de presença/ausência das mesmas em uma série de genomas da mesma espécie; ii) sua estrutura 3D e enovelamentos, mecanismos de transformação e resíduos envolvidos, cofatores, inibidores e parâmetros físicos, químicos e bioquímicos; iii) Análise do contexto genômico, essencialidade e taxa de crescimento de mutantes, iv) Perfis de expressão e de co-expressão. Para tanto, foi utilizado como modelo de estudo um conjunto de genomas correspondente a 33 linhagens de *E. coli* que inclui cepas patogênicas e não patogênicas.

## 1.2 Enzimas e sua classificação

As enzimas são capazes de catalisar reações bioquímicas e funcionam como catalisadores biológicos naturais. Elas diminuem a energia de ativação necessária para uma reação bioquímica, e aumentam a velocidade da reação enzimática. Em uma reação enzimática, as enzimas convertem o substrato em uma nova molécula chamada produto; durante este processo as enzimas podem requerer fatores adicionais para sua atividade catalítica, tais como cofatores, átomos de metais (por exemplo:  $\text{Fe}^{2+}$ ,  $\text{Zn}^{2+}$  e outros), ou outras moléculas que aumentam a afinidade da enzima pelo substrato.

A maioria dos processos metabólicos em um sistema biológico necessita de uma enzima. Possivelmente, a presença de enzimas no LUCA (último ancestral universal comum de todos os organismos) deve ter desempenhado um papel fundamental na evolução e irradiação da vida na Terra. As enzimas têm suas atividades biológicas definidas pelo tipo de transformação química realizada e pelo mecanismo através do qual esta transformação se processa.

As transformações químicas realizadas pelas enzimas são classificadas de acordo com as recomendações da “Nomenclature Committee of the International Union of Biochemistry” (NC-IUBMB, <http://www.chem.qmul.ac.uk/iubmb/enzyme/>), segundo o qual para cada atividade enzimática um nome recomendado e um número EC (“Enzyme Commission number”) são atribuídos, dependendo da reação catalisada pela enzima. Esta classificação é feita de forma hierárquica e cada enzima recebe um número composto de quatro dígitos; o primeiro dígito descreve a reação química geral catalisada pela enzima, ou seja, a classe a qual ela pertence; os dois números subsequentes têm significados diferentes dependendo da classe da enzima; o quarto dígito descreve a especificidade da reação, definindo o substrato/produto ou cofatores específicos usados (Quadro 1). Por exemplo, o EC 1.1.1.1 (álcool desidrogenase) representa as enzimas oxidoreduases (EC 1) que atuam nos grupamentos CH-OH de doadores (EC 1.1), com NAD ou NADP como aceptores (EC 1.1.1), tendo como substratos o álcool etílico ou hemiacetais (EC 1.1.1.1).

O número oficial de EC é organizado pela IUBMB e *International Union of Pure and Applied Chemistry* (IUPAC) é atribuído manualmente baseado em dados experimentais publicados sobre cada enzima. Estas informações compiladas fornecem uma biblioteca

valiosa sobre o conhecimento bioquímico atual proporcionando a classificação padrão e nomenclatura de enzimas (Kotera *et al.*, 2004, Kotera *et al.*, 2014).

Porém, nem toda enzima possui um EC; muitas reações do metabolismo secundário ainda não foram completamente elucidadas pelos métodos laboratoriais tradicionais (Kotera *et al.*, 2004). Contudo, o número EC é uma maneira inteligente de ligar a genômica e proteômica química, pois este número, além de representar a reação enzimática (informação química), é utilizado como identificador da enzima e dos genes que codificam estas enzimas (informação genômica). Desta forma, podemos ligar as duas informações para atribuir função a genes e proteínas identificadas em um determinado organismo (Kotera *et al.*, 2004).

**Quadro 1:** Classes enzimáticas e classificação segundo a União Internacional de Bioquímica

<b>Grupo de enzimas</b>	<b>Reação catalisada</b>	<b>Exemplos</b>
EC1: Oxidoreduases	Transferência de átomos ou moléculas de hidrogênio e oxigênio de um substrato para outro, ou associados às reações de oxidação.	Desidrogenases e oxidases
EC2: Transferases	Transferência de um grupo específico (fosfato, metil, etc) de um substrato para outro.	Transaminases e cinases
EC3: Hidrolases	Hidrólise de um substrato.	Enzimas digestivas
EC4: Liases	Remoção não hidrolítica de um grupo ou adição de um grupo ao substrato.	Aldolases
EC5: Isomerases	Mudança na forma molecular de um substrato.	Fumarases
EC6: Ligases	Junção de duas moléculas pela formação de novas ligações.	Sintetases

### 1.3 *Escherichia coli* e Pangenoma

A *E. coli* é uma das bactérias gram-negativas comensais do homem. Seu descobridor foi o alemão austríaco Theodor Escherichi em 1885. A bactéria é classificada dentro da subdivisão gama do Filo Proteobactéria, é um membro da Família Enterobacteriaceae, um dos organismos mais estudados pelo homem, e é amplamente utilizada como um vetor de clonagem em tecnologias do DNA recombinante. O ciclo de vida da bactéria entérica envolve a transição entre dois principais ambientes. O primeiro habitat é o trato gastrointestinal de mamíferos e o segundo é o solo, água e sedimentos (Gordon, Bauer, & Johnson, 2002).

O paradigma atual é que *E. coli* coloniza os humanos logo após o nascimento e se estabelece predominantemente como um anaeróbio facultativo no trato gastrointestinal pelo resto da vida do indivíduo (Bettelheim, 1976, Crossman *et al.*, 2010). A maioria das estirpes de *E. coli* são comensais, sendo inofensivas para o homem, porém algumas são patogênicas e podem ser transmitidas ao homem principalmente através do consumo de alimentos contaminados.

As estirpes causadoras de infecções intestinais podem ser divididas em seis tipos: *E. coli* enteroagregativa (EAEC), *E. coli* enteroinvasiva (EIEC), *E. coli* enteropatogênica (EPEC), *E. coli* enterohemorrágica (EHEC), *E. coli* de aderência efusiva (DAEC) e *E. coli* enterotoxigênica (ETEC) (Crossman *et al.*, 2010). As cepas patogênicas podem causar graves doenças, tais como diarreia hemorrágica, infecções extra-intestinais, septicemia e meningite, com complicações, muitas vezes graves que podem ser conduzidos para a morte; existem evidências de resistência ao tratamento de algumas cepas (Muniesa *et al.*, 2012, WHO, 2011; CDC, 2013). Surtos de doenças causadas por *E.coli* patogênicas têm ocorrido constantemente nos Estados Unidos (<http://www.cdc.gov/ecoli/>) (Rasko *et al.*, 2008a). Em maio de 2011, a linhagem O104: H4 causou 39 mortes e um surto de síndrome hemolítica urêmica (falência renal) e diarreia hemorrágica em 810 pacientes (Bielaszewska *et al.*, 2011)

As linhagens de *E. coli* podem diferir em mais de ¼ de seus genomas, como por exemplo, o genoma da *linhagem* K-12 apresenta uma diminuição de 1,4 megabases de DNA em comparação com o genoma da estirpe O157 (Dobrindt, 2005, Pallen & Wren,

2007). A classificação em patogênica ou comensal é baseada primariamente na presença ou ausência de certas regiões do DNA associadas com certos patótipos (Dozois & Curtiss, 1998)

Alguns estudos sugerem que as cepas patogênicas adquiriram elementos móveis codificando fatores de virulência através de transferência gênica horizontal (Escobar-Páramo *et al.*, 2004). Apesar disto, é ainda um desafio determinar e identificar se estas variações genômicas têm influência no metabolismo ou na adaptação destas bactérias em diferentes nichos; isto geralmente é observado através de fenótipos de resistência aos antibióticos e altas taxas de virulência.

As principais forças que promovem as variações genômicas intra-específicas são as mutações, o ganho e a perda de genes. O ganho de genes através de transferência horizontal parece ser a principal fonte de inovação e variação em bactérias (Pallen & Wren, 2007).

Outra fonte de inovação e variabilidade genética, é a transformação genética natural. Este processo consiste na captação de DNA exógeno (Plasmidial ou cromossômico) disperso no ambiente por bactérias e a incorporação deste material genético no genoma (Smith *et al.*, 1991).

A evolução de bactérias patogênicas de um ancestral não patogênico é marcada principalmente pela aquisição de grupos de genes de virulência e ilhas de patogenicidade através de conjugação de DNA por mecanismos de transferência gênica horizontal ou indiretamente via bacteriófagos que atuam como vetores (Pallen & Wren, 2007). As comparações do genoma entre distintas linhagens e a identificação de particularidades no aparato enzimático podem ajudar no entendimento destas variações fenotípicas de *E. coli*.

As ilhas de patogenicidades são blocos de genes (10-200kb) presentes em bactérias patogênicas que não estão presentes em linhagens não patogênicas destas bactérias (Schmidt & Hensel, 2004). Estas regiões têm como características um desvio composicional no uso de códons, que é marcado pelo aumento atípico no número de G+C em comparação com as demais regiões do genoma, sugerindo que estas regiões foram adquiridas por meio de transferência gênica horizontal. Outra característica, encontrada em regiões do DNA de bactérias com origem exógena, é a presença das seguintes marcas no genoma: genes de bacteriófagos, presença de elementos móveis e regiões de

inserção (IS) ou presença de genes que codificam para as enzimas transposases e integrases (Boyd *et al.*, 2009)

Estudos recentes têm demonstrado que a perda de genes de avirulência também é uma forma de aquisição de patogenicidade, pois, durante a evolução de microorganismos patogênicos cada bactéria adapta-se a seu estilo de vida através da redução da expressão de genes de avirulência ou pela sua deleção, um fenômeno conhecido como “buraco negro”; isto pode ser acompanhado por mutações dentro do gene ou supressão da expressão (Maurelli, 2007). Os genes de avirulência são definidos como genes cuja expressão em uma bactéria patogênica é incompatível com sua virulência (Maurelli, 2007).

Neste contexto, podemos afirmar que os estudos das variações e conservações nos genomas e proteomas de bactérias são de fundamental importância para a compreensão da biologia destes microrganismos, além de suas potenciais aplicações no desenvolvimento de novas drogas e antibióticos.

O advento do sequenciamento em larga escala tem possibilitado o sequenciamento do genoma de várias estirpes de bactérias (Tettelin, *et al.*, 2008), possibilitando o estudo das variações intra-específicas. Neste contexto, foi criado o termo pan-genoma. O pan-genoma denota o conjunto total de todos os genes presentes em uma espécie. Compreendendo o pan-genoma estão o genoma núcleo (*core*), que é o conjunto de genes comum a todos os membros de uma espécie e o genoma acessório, que contém genes que não estão presentes em todos os membros (Laing *et al.*, 2010). Além disto, o pan-genoma pode ser distinguido em dois tipos: pan-genoma aberto e pan-genoma fechado. O pan-genoma fechado denota que o sequenciamento de novos membros de uma espécie dificilmente irá render novos genes, o pan-genoma aberto indica que cada novo membro do sequenciamento irá render novos genes para aquela espécie (Mira *et al.*, 2010).

#### **1.4 Biologia de sistemas**

Os avanços tecnológicos nos últimos anos (século XXI) têm possibilitado o estudo dos organismos sobre diferentes escalas de complexidade biológica nas diferentes

ciências denominadas ômicas: genômica (estudo da sequência completa de um organismo), transcriptômica (conjunto de transcritos de RNA), proteômica (conjunto de proteínas de um organismo), epigenômica (conjunto de modificações no genoma e histonas através de metilações), interatômica (conjunto de interações de proteínas em um organismo). As tecnologias ômicas adotam uma visão holística das moléculas que constituem uma célula,

Dentre as técnicas de larga escala desenvolvidas nas últimas décadas podemos citar o sequenciamento em larga escala providos pelas tecnologias de nova geração (plataformas Roche/454, Illumina/Solexa, PacBio) (Liu *et al.*, 2012), microarranjos de DNA (Quackenbush, 2002) e proteínas (Templin *et al.*, 2002), sistemas duplos-híbridos e Chip-seq (Park, 2009).

Antes do advento destas tecnologias, os experimentos biológicos eram realizados em pequena escala visando estudar os componentes individuais da célula. Dentre as técnicas comumente utilizadas podemos citar: PCR (reação em cadeia da polimerase), macroarranjos e sequenciamento utilizando a tecnologia de Sanger. Tais estudos tinham como foco identificar os componentes individuais do sistema e a relação gene-fenótipo (características monogênicas).

Com o surgimento das tecnologias de larga escala, emergiu e tomou forma uma nova ciência denominada biologia de sistemas que têm como objetivo integrar todas as informações biológicas disponíveis obtidas de experimentos em larga escala (ciências ômicas) para obter uma visão holística do problema a ser estudado.

Com isto a visão reducionista está sendo substituída gradativamente por esta abordagem que visa estudar e compreender as múltiplas interações dentro de uma célula, tais como interação DNA e proteína, proteína e proteína, proteína e RNA e outras, e o fenótipo que emerge destas múltiplas interações nos sistemas biológicos durante uma condição fisiológica ou ambiental.

A explosão e disponibilidade de inúmeros dados biológicos na WEB gerados pelas recentes tecnologias de larga escala têm um valor inestimável para as pesquisas biológicas. Com os avanços nas tecnologias de seqüenciamento de DNA, a aquisição de dados genômicos se torna cada vez mais eficiente em termos de custos, e os conjuntos de dados estão se acumulando nos bancos de dados Web a uma taxa exponencial que supera nossa capacidade de analisar o seu significado biológico.

O entendimento destes dados requer métodos estatísticos e computacionais devido à complexidade de dados heterogêneos e volumosos. O estoque, processamento, consulta, análise e interpretação de tal incrível quantidade de dados é uma tarefa significativa que também têm muitos obstáculos e só pode ser executado com recursos computacionais de alto desempenho e com a bioinformática.

A bioinformática é a aplicação de ferramentas computacionais para predição, manutenção e interpretação dos dados biológicos (Swindells *et al.*, 2002). Dentre os bancos de dados biológicos podemos citar: GENE – NCBI (banco de dados do NCBI com informações sobre contexto genômico; [www.ncbi.nlm.nih.gov/gene](http://www.ncbi.nlm.nih.gov/gene)), PDB - “Protein Data Bank” (banco com estruturas tridimensionais de proteínas; [www.rcsb.org](http://www.rcsb.org)), KEGG - Enciclopédia de genes e genomas (banco com informações de atividades enzimáticas e vias metabólicas; [www.genome.jp/kegg](http://www.genome.jp/kegg)) e *STRING* (banco de dados com informações de interação proteína-proteína; [STRING -db.org/](http://STRING-db.org/)).

A bioinformática e as análises evolutivas geralmente começam com a comparação entre sequências de genes e proteínas de diferentes organismos através do alinhamento entre elas (alinhamento par-par) visando inferir a relação entre estrutura/função e posteriormente história evolutiva para um ou mais genes ou proteínas. A função de uma proteína pode ser descoberta usando a informação de uma ou mais proteínas conhecidas através de um alinhamento de sequências. Porém, o método mais comumente usado e mais acurado é para inferir a função de uma proteína é a modelagem comparativa (Lam et al., 2017).

Geralmente, todos os métodos de seleção de modelos envolvem a pesquisa de estruturas de proteínas modelo a partir do PDB. A identidade de sequência global entre a consulta e os modelos foi amplamente utilizada como o principal critério em um processo de pesquisa usando o BLAST (Altschul et al., 1990). BLAST alinha duas sequências com base em uma matriz de substituição, contendo um esquema de pontuação (score) quando dois resíduos de aminoácidos são alinhados. Esta matriz de substituição captura a probabilidade com a qual um resíduo de aminoácido específico sofra uma mutação para outro resíduo substituído durante um longo período evolutivo (Lam et al., 2017).

A modelagem comparativa de proteínas geralmente produz um bom modelo tridimensional se uma sequência de proteína homóloga possui uma identidade global de

≥30%. Contudo uma vez que a identidade entre as sequências diminui abaixo de 30% (a 'twilight zone'), a qualidade do modelo deteriora rapidamente (Baker & Sali, 2001).

Com a quantidade massiva dos dados biológicos disponíveis na WEB inúmeras perguntas biológicas podem ser respondidas com a obtenção, processamento e análises estatísticas destes dados.

Os estudos de biologia de sistemas podem ser utilizados na identificação de marcadores moleculares para diagnósticos de doenças genéticas, no estudo das respostas celulares ou determinantes genéticos que influenciam na metabolização de um fármaco ou de uma vacina, na identificação e proposição de alvos moleculares para fármacos em patógenos e outras aplicações.

Nos próximos anos, possivelmente com os conhecimentos adquiridos utilizando uma visão de biologia de sistemas poderemos simular ou até prever possíveis efeitos de drogas ou outros agentes em uma célula e identificar os componentes e as propriedades essenciais que regem as múltiplas interações em um sistema biológico.

## **2 Objetivo geral**

O presente trabalho tem como objetivo verificar se as enzimas análogas intragenômicas representam casos de redundância funcional ou oferecem flexibilidade metabólica aos organismos, e analisar o papel biológico destas enzimas com ênfase em sua associação com mecanismos de patogenicidade.

### **2.1 Objetivos Específicos**

- 1.- Identificar os casos de analogia intragenômica em um conjunto de linhagens patogênicas e não patogênicas de *E. coli*;
- 2.- Examinar se essas enzimas representam casos de redundância funcional;
3. Estabelecer, com base nos dados gerados e na análise da literatura, o papel biológico de cada forma estrutural alternativa.

### 3 Metodologia

#### 3.1 Importação de dados do KEGG e predição computacional das formas análogas

O conjunto de dados correspondente a todas as proteínas computacionalmente preditas ou experimentalmente verificadas representantes de todas as atividades enzimáticas descritas até junho de 2011 foi obtido (em formato fasta) do banco de dados KEGG (Kanehisa & Goto, 2000) versão 58.1 (<http://www.genome.jp/kegg>). Esse banco de dados contém uma coleção de mapas manualmente elaborados, representando o conhecimento atual sobre redes de interação molecular em processos biológicos (vias bioquímicas). Tal versão possuía 1.535 organismos, divididos em: eucariotos (142), Eubactérias (1.286) e Archea (107). As sequências proteicas representantes de todas as atividades enzimáticas descritas para os genomas destes organismos foram analisadas através de um fluxo de trabalho pipeline implementado na ferramenta AnEnPi (*Analogous Enzyme Pipeline*) (Otto *et al.*, 2008) (<http://bioinfo.pdtis.fiocruz.br/AnEnPi/>), que permite a predição computacional de enzimas análogas funcionais em grupos de sequências proteicas. O algoritmo empregado neste programa consiste nos seguintes passos:

(i) Remoção de sequências muito pequenas, menores do que 100 aminoácidos, do conjunto de dados; tais sequências podem representar artefatos oriundos dos processos de anotação genômica;

(ii) Comparação entre todas as sequências proteicas, separadamente por atividade enzimática, com o programa blastp do pacote de programas de busca por similaridade local BLAST (Altschul *et al.*, 1997);

(iii) Agrupamento das sequências comparadas, separadamente por atividade enzimática, de acordo com o grau de similaridade entre elas, como descrita por Galperin e colaboradores (1998). Os resultados obtidos nesta comparação (pares alinhados) são transformados em um grafo no qual cada enzima é representada por um nó. Se duas enzimas pertencentes a uma mesma atividade enzimática, classificadas com o mesmo número EC, obtêm uma pontuação (*score*) maior ou igual a 120 quando comparadas com Blastp, os dois nós somos conectados por uma aresta. Todas as sequências conectadas

no grafo são incluídas no mesmo grupo, sendo consideradas presumidamente homólogas. Sequências não conectadas, ou seja, pertencentes a grupos distintos de uma mesma atividade enzimática, são consideradas presumidamente análogas. O número de sub-grafos representa o número de eventos de origem independente em cada atividade enzimática ou, em outras palavras, o número de vezes em que determinada atividade enzimática surgiu *de novo* durante a evolução. Conforme descrito na introdução deste trabalho, o limiar de similaridade utilizado na formação dos grupos é baseado em uma observação experimental importante: enzimas que comprovadamente compartilham a mesma atividade enzimática e que apresentam estruturas tridimensionais significativamente distintas, medido através de alinhamentos estruturais, quando foram comparadas ao nível de suas estruturas primárias (sequências de aminoácidos) com o programa blastp obtiveram pontuação abaixo de 120 (Galperin *et al.*, 1998).

### **3.2 Identificação de análogos intragenômicos em *E. coli***

Os dados obtidos da etapa de agrupamento foram filtrados através de *scripts* desenvolvidos em Perl, e foi obtida uma tabela contendo a quantidade de grupos gerados para cada atividade enzimática de todos os organismos contidos em nosso banco de dados local. Posteriormente, foi gerada uma tabela exclusivamente para os casos de analogia intragenômica nas linhagens de *E. coli* estudadas.

Ao mesmo tempo, foi realizada uma segunda etapa de identificação de NISEs denominada inferência de função, esta etapa consiste na re-anotação dos genomas de *E. coli* e teve como finalidade unificar as anotações dos genomas analisados, contribuindo assim para aumentar o repertório de casos de analogia identificados, principalmente para os casos de proteínas do KEGG anotadas como hipotéticas ou de função não conhecida, pois sem isto estas proteínas são perdidas na etapa de agrupamento. Inicialmente, os proteomas das linhagens de *E. coli* foram obtidos no banco de dados RefSeq: (<ftp://ftp.ncbi.nlm.nih.gov/genomes/>), em seguida a inferência funcional foi realizada com a ferramenta BLASTP; o proteoma de cada bactéria foi comparado contra cada grupo de sequências geradas na etapa de agrupamento utilizando um valor de corte (*E-value*) de  $e^{-20}$  conforme Gomes *et al.*, 2011. No final desta etapa, nós obtivemos uma segunda lista, contendo os casos de analogia totais, ou seja, a soma das duas listas de análogos

intragenômicos obtidos em *E. coli*, retirando os casos redundantes. As enzimas heteromultiméricas, que são enzimas constituídas por mais de uma subunidade enzimática, codificadas por diferentes genes, foram excluídas das análises. Para realizar esta etapa, as sequências anotadas com o termo *subunit* ou *heavy chain* foram excluídas e utilizamos também informações da organização estrutural das enzimas no banco de dados PDBSum ([www.ebi.ac.uk/pdbsum/](http://www.ebi.ac.uk/pdbsum/)) e PDBe ([www.ebi.ac.uk/pdbe/](http://www.ebi.ac.uk/pdbe/)). Para a reconstrução de vias metabólicas, os mapas do metabolismo de *E. coli* foram gerados através da ferramenta KEGGMAPPER (<http://www.genome.jp/kegg/pathway.html>), inserindo as listas de ECs (atividades enzimáticas) obtidas na etapa de inferência funcional.

### **3.3 Classificação estrutural dos enovelamentos estruturais e superfamílias das proteínas preditas presentes nos proteomas das linhagens de *E. coli***

As sequências protéicas identificadas em *E. coli* K12 MG1655 agrupadas em um mesmo *cluster* e as sequências agrupadas em *clusters* diferentes foram re-anotadas utilizando o banco de dados UNIPROTKB/SWISPROT (<http://www.uniprot.org/>) e ENZYME (<http://enzyme.expasy.org/>). As sequências primárias das proteínas (em formato *fasta*) de *E. coli* alocadas nos grupos de homólogo foram recuperadas através de seus identificadores com o auxílio de scripts em Perl. Através do banco de dados SUPERFAMILY ([supfam.cs.bris.ac.uk/](http://supfam.cs.bris.ac.uk/)) e SCOP ([scop.mrc-lmb.cam.ac.uk/shop/](http://scop.mrc-lmb.cam.ac.uk/shop/)), fizemos a predição dos enovelamentos estruturais destas sequências.

Esta etapa é crucial para validação dos casos de analogia identificados na predição computacional. As enzimas com um mesmo número de EC com estruturas tridimensionais (3D) classificadas em distintas Superfamílias e distintas conformações nos bancos de dados SUPERFAMILY e SCOOP, respectivamente, são consideradas verdadeiros análogos funcionais. E as enzimas homólogas são as que foram classificadas nas mesmas superfamílias e com as mesmas conformações estruturais, embora algumas vezes estas sequências tenham sido alocadas em grupos diferentes durante a etapa de agrupamento.

Na base de dados SUPERFAMILY, é possível comparar uma sequência com os modelos HMMs (*Hidden Markov Model*) obtidos de todas as estruturas 3D depositadas no banco de dados PDB. Com isto, mesmo na ausência de um modelo 3D para uma proteína, é possível prever seu enovelamento e com isto refinar a identificação de casos

de analogia entre as sequências. Posteriormente, as enzimas análogas identificadas em *E. coli* foram alinhadas no PDB através da ferramenta BLAST, com o intuito de obter a sequência mais similar neste banco de dados, visando posteriormente a construção do modelo tridimensional destas proteínas e usar a informação do PDB para obter dados sobre a proteína que identificamos. Para as análises de similaridade de sequências foram utilizados os seguintes programas: Needle ([http://www.ebi.ac.uk/Tools/psa/emboss\\_needle/nucleotide.html](http://www.ebi.ac.uk/Tools/psa/emboss_needle/nucleotide.html)), Water ([http://www.ebi.ac.uk/Tools/psa/emboss\\_water/](http://www.ebi.ac.uk/Tools/psa/emboss_water/)) e Muscle (Edgar, 2004), implementado no programa Seaview (Galtier *et al.*, 1996)

### 3.4 Modelagem por homologia e análise estrutural comparativa 3D

A etapa de modelagem comparativa por homologia foi realizada em colaboração com o Dr. *Floriano Paes* Silva Júnior do Laboratório de Bioquímica de Proteínas e Peptídeos, Instituto Oswaldo Cruz, Fiocruz e MsC. Paulo Vinícius Sanches Daltro de Carvalho. A realização desta etapa teve como objetivo obter o modelo 3D de algumas enzimas análogas para uma análise mais acurada dos distintos enovelamentos, sítio ativo e mecanismos de transformação química empregados por estas enzimas. A modelagem das estruturas tridimensionais foi executada pelo método de satisfação de restrições espaciais implementados no programa MODELLER 9v10 ([https://salilab.org/modeller/old\\_releases.html](https://salilab.org/modeller/old_releases.html)). Os modelos foram gerados usando as coordenadas atômicas dos moldes obtidos no banco de dados PDB (*Tabela S2*). Os modelos das estruturas 3D de algumas sequências primárias foram baseados em 5 moldes visando obter um modelo de melhor qualidade estéreo-química. Para cada estrutura foram gerados cinco modelos. A seleção do modelo foi executada pela função de minimização de energia normalizada DOPE score GA341 (Eramian *et al.*, 2006). O programa PROCHECK (Laskowski *et al.*, 1993) foi usado para verificar e analisar a qualidade estéreo-química e qualidade global da estrutura tridimensional (3D) através dos gráficos de Ramachandran, permitindo a análise da geometria dos resíduos de aminoácidos e geometria total da estrutura. Todos os modelos selecionados para análises posteriores possuem mais de 90% dos resíduos nas regiões mais favoráveis e permissivas do gráfico de “Ramachandran”. Visando executar a análise de uma região

particular, entre as estruturas dos análogos ou NISES, primeiro foi obtida e transferida à coordenada de um ligante particular através dos dados obtidos da estrutura 3D dos homólogos no PDB referente a cada análogo. Na etapa final, a minimização do complexo de energia (*protein-binding*) foi realizada com a interface do *programa* Maestro ([www.schrodinger.com/Maestro/](http://www.schrodinger.com/Maestro/)) usando um campo de força OLPS2005. A visualização e análise das estruturas foi realizada utilizando o *programa* PyMOL ([www.pymol.org/](http://www.pymol.org/)) e Maestro; os detalhes do sítio ativo próximos aos ligantes foram mostrados em uma distância de 5 Å (angstrom). Os resíduos dos sítios catalíticos e alostéricos presentes nas estruturas 3D foram identificados usando o banco de dados *Catalytic Site Atlas* (<http://www.ebi.ac.uk/thornton-srv/databases/CSA/>) e através de informações obtidas na literatura.

### **3.5- Identificação e análise de cofatores, substratos, parâmetros cinéticos, inibidores, localização celular, parâmetros físicos químicos e bioquímicos.**

As informações sobre os parâmetros bioquímicos: cofatores e substratos utilizados por cada análogo, parâmetros cinéticos (velocidade da reação), afinidade enzimática (KM: constante Michaelis), inibidores enzimáticos e localização celular foram obtidas a partir de múltiplas fontes de dados; foram utilizados os bancos de dados BRENDA ([www.brenda-enzymes.info/](http://www.brenda-enzymes.info/)), EcoCyc ([ecocyc.org/](http://ecocyc.org/)) e UNIPROTKB ([www.uniprot.org](http://www.uniprot.org/)). Os custos metabólicos, ou seja, o total de ATP utilizado para a síntese das formas análogas foram obtidos da literatura (Smith & Chapman, 2010). Os parâmetros físicos e químicos foram determinados com a ferramenta ProtParam (<http://web.expasy.org/protparam/>), onde foram calculados os seguintes parâmetros: tamanho da proteína, total de resíduos positivamente e negativamente carregados (pH), composição de aminoácidos, total de átomos, índice de estabilidade e grande média de hidropaticidade (GRAVY).

### **3.6 Análises de rede de interação dos análogos intragenômicos em *E. coli***

Os dados da rede de interação dos análogos intragenômicos foram obtidos no banco de dados *STRING* (*STRING-db.org/*). As evidências para associação do banco de dados *STRING* são baseadas em múltiplas fontes (von Mering *et al.*, 2005), incluindo: i) Métodos de contexto genômico (vizinhança, fusão gênica e co-ocorrência); ii) co-expressão de genes; iii) dados experimentais de interação proteína-proteína importado dos banco de dados BIND (<http://bind.ca>), DIP (<http://dip.doe-mbi.ucla.edu/dip/Main.cgi>), GRID (<http://thebiogrid.org/>), HPRD (<http://www.hprd.org/>), MINT (<http://mint.bio.uniroma2.it/mint/>); iv) dados curados dos bancos de dados KEGG ([www.genome.jp/kegg/](http://www.genome.jp/kegg/)), REACTOME ([www.reactome.org/](http://www.reactome.org/)), MIPS (catálogo de complexos protéicos, [www.mips.helmholtz-muenchen.de/proj/ppi](http://www.mips.helmholtz-muenchen.de/proj/ppi)) , BIOCARTA (catálogo de vias metabólicas em *Homo sapiens*, [www.biocarta.com/](http://www.biocarta.com/)) e STKE (catálogo de vias de sinais de transdução em *Homo sapiens*, <http://www.usc.edu/libraries/>); v) genes co-mencionados em resumos de artigos da base de dados Medline (<http://www.nlm.nih.gov/bsd/pmresources.html>), parágrafos SGD (<http://www.yeastgenome.org>) e monografias OMIM (<http://www.ncbi.nlm.nih.gov/omim>). Para todas estas fontes, o banco de dados *STRING* transfere automaticamente as informações entre os organismos modelos quando possível usando um processo automatizado. Os dados obtidos no banco de dados *STRING* foram analisados no programa *Cytoscape* (versão 2.8, [www.Cytoscape.org/](http://www.Cytoscape.org/)) com o subprograma Network Analysis para calcular os valores de grau de interação e *betweenness* dos nós que constituem a rede biológica. Posteriormente, a rede de interação (conjunto de proteínas) de cada análogo foi classificada funcionalmente através do banco de dados David Bioinformatics Resouces 6.7 (<http://david.abcc.ncifcrf.gov/>), utilizando os seguintes parâmetros: severidade de classificação alta, testes estatísticos FDR, Benjamini-Hochberg (Benjamini & Hochberg, 1995), *Bonferroni* (Hochberg & Yosef, 1988).

### 3.7 Análises de expressão e co-expressão gênica

Para as análises de expressão gênica, os dados de micro-arranjos (GSE1121, GSE4556, GSE4778, GSE13418) foram obtidos da base de dados Gene Expression Omnibus (GEO) (<http://www.ncbi.nlm.nih.gov/geo/>) e processados. O processamento consiste nas etapas de normalização, anotação dos genes e identificação dos genes diferencialmente expressos. Inicialmente os dados foram transformados na base 2 da escala log e depois normalizados pelo método RMA (*Robust Multi-Array Analysis Average*), para cada gene e amostra um p-value foi calculado usando o teste de correções múltiplas FDR (*False discovered rates*, Benjamini & Hochberg, 1995). Os resultados foram filtrados com base na significância estatística utilizando um p-valor de 0.05 (95% de confiança) e “Fold Change” > 1”. Todas as análises foram executadas no *programa* R ([www.r-project.org/](http://www.r-project.org/)) utilizando bibliotecas da plataforma Bioconductor ([www.bioconductor.org/](http://www.bioconductor.org/)). Para a obtenção e análise dos dados, as seguintes bibliotecas foram utilizadas: GEOquery, RMA (*Robust Multi-Array Analysis Average*), simpleaffy, AnnotationDBi, RColorBrewer e Limma. Os dados foram re-analisados utilizando o *programa* GENEVESTIGATOR (<https://www.genevestigator.com/gv/>) e os valores de expressão gênica foram obtidos na forma de mapas de calor e gráficos em linhas. O padrão de co-expressão em uma mesma condição foi analisado no *programa* TMEV ([www.tm4.org/mev](http://www.tm4.org/mev)). usando o algoritmo agrupamento hierárquico K-means dividindo a amostra em 10 partições.

### 3.8 Análise do contexto genômico

Os genes e operons que flanqueiam o operon que contendo as sequencias que codificam as enzimas análogas intragenômicas foram analisados no banco de dados Gene (NCBI: [www.ncbi.nlm.nih.gov/gene](http://www.ncbi.nlm.nih.gov/gene)) e EcoCyc (<http://ecocyc.org/>). As figuras de contexto genômico foram geradas com a utilização da ferramenta Gene Context Tool (GeConT) (<http://bioinfo.ibt.unam.mx/gecont>.) ou obtidas diretamente do EcoCyc; a localização esquemática dos genes no genoma foi obtida com a ferramenta Circos ([circos.circos.ca/software/download/circos/](http://circos.circos.ca/software/download/circos/)).

### 3.9 Análise dos fenótipos de mutantes em *Escherichia coli*, essencialidade, resistência para antibióticos e *fitness*

Os dados de essencialidade foram obtidos nos bancos de dados PEC ([www.shigen.nig.ac.jp/ecoli/PEC/index.jsp](http://www.shigen.nig.ac.jp/ecoli/PEC/index.jsp)) e através de inspeção manual na literatura. Os dados da taxa de sobrevivência e adaptação (*fitness*) dos mutantes em *E. coli* foram obtidos na base de dados PortEco (<http://www.porteco.org/phenotypes/>) (Nichols *et al.*, 2011). O banco de dados Antibiotic Resistance Genes Database (<http://ardb.cbcb.umd.edu>) foi utilizado para analisar se as enzimas identificadas possuem similaridade com proteínas relacionadas à resistência à drogas e o banco de dados Drug bank ([www.drugbank.ca/databases](http://www.drugbank.ca/databases)) foi utilizado para avaliar se as enzimas identificadas estão sendo estudadas como alvos potenciais para drogas. Também examinamos a presença de formas estruturais enzimáticas com casos de analogia em plasmídeos através do banco de dados de plasmídeos (<http://www.ebi.ac.uk/genomes/plasmid.html>).

#### 3.10 – Mineração na Literatura

O papel biológico dos genes que codificam as enzimas análogas foi consultado no banco de dados EcoCyc (<https://ecocyc.org/>), um banco de dados que descreve o genoma e a maquinaria bioquímica de *E. coli* K-12 MG1655, é uma referência eletrônica para biólogos que incluem páginas de informação sobre cada gene de *E. coli*, metabólito, reação, operon e via metabólica. O banco de dados também inclui informação de essencialidade sobre cada gene, condições de nutrientes que suportam ou não o crescimento de *E. coli*, e uma página de referências sobre cada gene de *E. coli*. Os artigos listados no banco de dados EcoCyc foram utilizados para mineração na literatura dos papéis biológicos de cada gene.

## 4 RESULTADOS

Todos os resultados obtidos a partir da aplicação destes métodos para cada atividade enzimática foram utilizados para avaliar hipótese inicial, verificando se as enzimas isofuncionais não-homólogas representam casos de redundância funcional ou, alternativamente, desempenham distintos papéis biológicos, utilizando para tal demonstração os genomas completos de 33 linhagens de *E. coli* (Tabela 1). Apenas alguns destes resultados serão mostrados neste trabalho. Os resultados completos estão arquivados em um CD em anexo.

Tabela 1: Linhagens de *E. coli* cujos genomas foram utilizados neste trabalho.

Código KEGG	Linhagens de <i>Escherichia coli</i>	Patogenicidade*	ID Banco de dados GENOME/NCBI
Ece	O157 EDL933(EHEC)	Patogênica	GCA_000006665.1_ASM666v1
Ecs	O157: H7 Sakai (EHEC)	Patogênica	GCF_000008865.2_ASM886v2
Ecf	O157: H7 EC4115	Patogênica	GCA_000021125.1_ASM2112v1
Eoh	O103: H2 12009 (EHEC)	Patogênica	GCA_000010745.1_ASM1074v1
Eoi	O111: H- 11128 (EHEC)	Patogênica	GCA_000010765.1_ASM1076v1
Eoj	O26: H11 11368 (EHEC)	Patogênica	GCA_000091005.1_ASM9100v1
Etw	O157: H7 TW14359	Patogênica	GCA_000022225.1_ASM2222v1
Eok	O55: H7 CB9615	Patogênica	GCA_000025165.1_ASM2516v1
Ecg	O127: H6 E2348/69	Patogênica	GCA_000026545.1_ASM2654v1
Eck	55989 (EAEC)	Patogênica	GCA_000026245.1_ASM2624v1
Ecv	O1: K1: H7 (APEC)	Patogênica	GCA_000014845.1_ASM1484v1
Ecw	O139: H28 E24377A	Patogênica	GCA_000017745.1_ASM1774v1
EcZ	O45: K1: H7 S88	Patogênica	GCA_000026285.2_ASM2628v2
Eci	O18: K1: H7 UTI89	Patogênica	GCA_000013265.1_ASM1326v1
Ecp	O6: K15: H31 536	Patogênica	GCA_000013305.1_ASM1330v1
Ecc	O6: K2: H1 CFT073	Patogênica	GCA_000007445.1_ASM744v1
Eum	O17: K52: H18 UMN026	Patogênica	GCF_000026325.1_ASM2632v1
Eco	K-12 MG1655	Não patogênica	GCF_000005845.2_ASM584v2
Ecd	K-12 DH10B	Não patogênica	GCA_000019425.1_ASM1942v1
Ebd	BL21-Gold(DE3)pLysS AG	Não patogênica	GCA_000023665.1_ASM2366v1
Ebr	B REL606	Não patogênica	GCA_000017985.1_ASM1798v1
Ebw	K-12 MC4100(MuLac) BW2952	Não patogênica	GCA_000022345.1_ASM2234v1
Ecj	K-12 W3110	Não patogênica	GCA_000010245.1_ASM1024v1
Ecl	C ATCC 8739	Não patogênica	GCA_000019385.1_ASM1938v1
Ecq	O81 ED1a (comensal)	Não patogênica	GCA_000026305.1_ASM2630v1
Ecr	O8 IAI1 (comensal)	Não patogênica	GCA_000026265.1_ASM2626v1
Ect	O7: K1 IAI39 (ExPEC)	Não patogênica	GCF_000026345.1_ASM2634v1
EcX	O9 HS (comensal)	Não patogênica	GCA_000017765.1_ASM1776v1
Ecy	O152:H28 SE11 (comensal)	Não patogênica	GCA_000010385.1_ASM1038v1
Ecm	SMS-3-5 (ambiental)	Não patogênica	GCA_000019645.1_ASM1964v1
Eso	O104:H4 2009EL-2071	Não patogênica	GCA_000299475.1_ASM29947v1
Esm	O104:H4 2009EL-2050	Não patogênica	GCA_000299255.1_ASM29925v1
esl	O104:H4 2011C-3493	Não patogênica	GCF_000299455.1_ASM29945v1

\*Informação obtida do banco de dados Genome (<https://www.ncbi.nlm.nih.gov/genome/>) do NCBI.

#### 4.1 Identificação de análogos funcionais em *E. coli*

Os casos de analogia funcional identificados neste trabalho foram preditos através de módulos do programa AnEnPi (Otto *et al.*, 2008). Para isto, nós obtivemos o conjunto de todas as atividades enzimáticas (representados por 1 milhão de diferentes atividades enzimáticas ou ECs completos, ou seja, contendo os 4 dígitos iniciais) disponíveis no banco de dados KEGG. Através de agrupamento (*clustering*), nós identificamos 815 atividades enzimáticas contendo mais do que um grupo. Estes grupos foram utilizados para a etapa de re-anotação dos genomas das linhagens de *Escherichia coli* (Tabela 1). Após a etapa de re-anotação, as sequências alocadas em distintos grupos e com a mesma atividade enzimática foram consideradas potencialmente análogas e as sequências agrupadas no mesmo grupo foram consideradas potencialmente homólogas. Inicialmente, foi identificado em torno de 70 potenciais atividades enzimáticas com casos de analogia nos genomas das linhagens de *E. coli* estudadas. Contudo, para refinar a predição de casos verdadeiros de analogia funcional, foram aplicados os seguintes filtros: i) remoção das enzimas heteromultiméricas (enzimas anotadas com o termo “*subunit*” e/ou “*heavy chain*”); ii) remoção de enzimas que possuem qualquer enovelamento (*fold*) em comum e classificada na mesma superfamília no banco de dados SUPERFAMILY(<http://supfam.cs.bris.ac.uk/SUPERFAMILY/>).

As enzimas heteromultiméricas são constituídas por diferentes subunidades codificadas por genes diferentes com distintas funções e origens evolutivas. Observamos que a anotação de uma determinada atividade enzimática (presente, por exemplo, na base de dados BRENDA - <http://www.brenda-enzymes.org/>), anotação esta que pode ser específica para uma subunidade particular de uma enzima heteromultimérica ou para esta como um todo, é normalmente herdada por todas as subunidades que compõem aquela enzima. Neste caso, falsos casos de analogia irão aparecer para os casos de enzimas heteromultiméricas, cuja atividade enzimática seja dependente da presença de mais de uma subunidade. De fato, observamos que atividades enzimáticas contendo enzimas heteromultiméricas possuem comumente um número bem maior de grupos em comparação com as demais atividades enzimáticas onde casos de analogia foram identificados, e a anotação das sequências presentes nestes grupos geralmente possuem o termo *subunit* ou *chain*. Conseqüentemente, como a presença das enzimas

heteromultiméricas poderia ser um fator confundidor na identificação de casos de analogia, as mesmas foram retiradas desta análise.

Através da análise dos dados obtidos a partir do banco de dados SUPERFAMILY, referentes aos padrões de enovelamento e superfamília, identificamos enzimas que foram alocadas em distintos grupos, ou seja, sem similaridade detectável entre suas estruturas primárias, mas com conservação entre seus padrões de enovelamento (*Tabela S1*). Isto indica que estas enzimas, ao invés de representarem um caso de analogia, eram na verdade homólogos distantes, sendo, portanto, classificadas como tal. Em um cenário alternativo, a similaridade entre suas estruturas tridimensionais poderia ser considerada como o resultado de evolução convergente, mas encontramos ao menos um trabalho que sugere que tais casos seriam raros (Doolittle, 1994). Um melhor exame da literatura e dos dados obtidos neste trabalho pode servir para obtermos dados mais precisos em relação a este aspecto.

Nós identificamos casos de NISES em cinco das seis classes enzimáticas: nas Oxidoredutases (EC 1 – nove casos), Transferases (EC 2 – quatro casos), Hidrolases (EC 3 – doze casos), Liases (EC 4 – três casos) e Isomerases (EC 5 – quatro casos); nenhum caso de analogia funcional foi identificado na classe das Ligases (EC 6) (*Tabela S2*). Com a metodologia empregada neste trabalho, conseguimos validar 33 casos de analogia em *E. coli* K12 (*Tabela 2*); destes, quatro casos foram considerados como NISES preditas, pois não foi possível determinar o enovelamento destas enzimas. Nossa abordagem permitiu adicionalmente anotar (atribuir uma possível atividade enzimática) algumas proteínas de *E. coli* K12 não caracterizadas ou não anotadas no UNIPROTKB e no KEGG (*Tabela S2*). Analisamos o padrão de presença e ausência de analogia intragenômica no contexto do pan-genoma constituído pelos genes presentes em 33 linhagens de *E. coli*. As enzimas presentes em todas as linhagens de *E. coli* foram caracterizadas como enzimas do genoma núcleo (*core genome*), e as não presentes em todas as linhagens foram classificadas como enzimas do genoma dispensável ou acessório (*accessory genome*). A *Tabela 2* mostra os 28 casos encontrados de atividades enzimáticas com formas análogas presentes no genoma núcleo; a maioria destas enzimas participa em vias metabólicas importantes para a manutenção do metabolismo e para a sobrevivência de *E. coli*. No genoma acessório, identificamos 5 atividades enzimáticas com casos de analogia, incluindo proteínas envolvidas com a resistência às drogas (*Tabela 2*). O padrão de presença e ausência das atividades enzimáticas com casos de analogia intragenômica

e suas respectivas formas enzimáticas são mostrados na Tabela S3. A Figura 1 mostra, de forma esquemática, a localização das enzimas análogas no genoma de *E. coli*.

Tabela 2: Enzimas análogas intragenômicas em *E. coli*

Número EC	Nome da enzima	Genes	Nº formas	Genoma núcleo/acessório	Via metabólica
1.1.1.1	Álcool desidrogenase	<i>adhE</i>	2	Núcleo	Glicólise e Gluconeogênese (eco00010)*
		<i>adhP</i>			
1.1.1.28	D-lactato desidrogenase	<i>Dld</i>	2	Núcleo	Metabolismo do piruvato (eco00620)
		<i>ldhA</i>			
1.11.1.6	Catalase	<i>Kate</i>	2	Núcleo	Metabolismo do glioxilato e dicarboxilato (eco00630); Metabolismo do triptofano (eco00380)
		<i>katG</i>			
1.1.1.20 5	Inosina-5'-monofosfato desidrogenase	<i>guaB</i>	2	Núcleo	Metabolismo de purinas (eco00230)
		<i>yjD<sup>d</sup></i>			
		<i>ytfL<sup>d</sup></i>			
1.15.1.1	Superóxido dismutase	<i>sodC</i>	2	Núcleo	N/A
		<i>sodA<sup>d</sup></i>			
		<i>sodB<sup>d</sup></i>			
1.5.1.38	FMN reductase (NADPH)	<i>Fre</i>	2	Núcleo	Metabolismo de riboflavinas (eco00740)
		<i>ssuE</i>			
1.5.1.3	Di-hidrofolato redutase	<i>folA<sup>e</sup></i>	2	Núcleo	Biosíntese de folato (eco00790)*
		<i>folM</i>			
1.97.1.9	Selenato redutase	<i>ygfK</i>	2	Núcleo	Metabolismo de seleno-compostos (eco00450)
		<i>yntE<sup>d</sup></i>			
		<i>yntF<sup>d</sup></i>			
2.1.2.2	Fosforibosil glicinamida formil transferase	<i>purN</i>	2	Núcleo	Metabolismo de purinas (eco00230)*
		<i>purT</i>			
2.3.1.15	Glicerol-3-fosfato 1-O-aciltransferase	<i>plsX</i>	2	Núcleo	Metabolismo de glicerofosfolipídeos (eco00564) e glicerolipídeos (eco00561)
		<i>plsB<sup>e</sup></i>			
		<i>plsY</i>			
2.3.1.18	Galactosideo O-acetiltransferase	<i>Laca</i>	2	Núcleo	N/D
		<i>Yia</i>			
2.5.1.17	Ácido Cob(II)irínico a,c-diamide adenosiltransferase	<i>Btur</i>	2	Núcleo	Metabolismo de porfirina e clorofila (eco00860)
		<i>euT</i>			
2.7.1.11	6-fosfofrutoquinase	<i>pfkA</i>	2	Núcleo	Metabolismo de frutose e manose (eco00051), biosíntese de aminoácidos (eco01230), glicólise e gluconeogênese (eco00010)*
		<i>pfkB</i>			
2.7.1.4	Frutoquinase	<i>O3M_07610</i>	3	Núcleo	N/D
		<i>O3M_19525</i>			
		<i>O3M_04850</i>			
3.1.1.5	Lisofosfolipase	<i>tesA</i>	3	Núcleo	Biosíntese de ácidos graxos insaturados (eco01040), glicerofosfolipídeos (eco00564)
		<i>rssA</i>			
		<i>pldB</i>			
3.1.22.4	Endodeoxiribonuclease ( <i>crossover junction</i> )	<i>rusA<sup>e</sup></i>	2	Núcleo	Recombinação homóloga (eco03440)
		<i>ruvC</i>			

3.1.3.1	Fosfatase alcalina	<i>phoA</i> <i>ybaA</i>	2	Núcleo	Sistema de dois-componentes (eco02020), biosíntese de folato (eco00790)*
3.1.3.2	Fosfatase ácida	<i>Appa</i> <i>Afa</i>	2	Núcleo	Metabolismo de riboflavinas (eco00740)*
3.2.1.17	Lisozima	<i>rrrD/ydfQ</i> <i>yegX</i> <i>Chia</i> <i>R</i>	4	Acessório	N/D
3.1.3.27	Fosfatidil-glicerofosfatase	<i>pgpB</i> <i>pgpA</i>	2	Núcleo	Metabolismo de glicerofosfolipídeos (eco00564)
3.1.3.5	5'-nucleotidase	<i>surE</i> <i>ushA</i> <i>yfbR</i> <i>yjiG</i>	4	Núcleo	Metabolismo de pirimidinas (eco00240)*
3.2.1.86	6-fosfo-beta-glicosidase	<i>ascB</i> <i>bgIB<sup>d</sup></i> <i>bgIA<sup>d</sup></i>	2	Núcleo	Glicólise e gluconeogênese (eco00010)
3.5.2.6	Beta-lactamase	<i>ampC</i> <i>ybeQ</i>	2	Acessório	Biosíntese de penicilina e cefalosporina (ec00311)*
3.6.1.1	Difosfatase inorgânica	<i>Ppa</i> <i>yiaD</i>	2	Acessório	Fosforilação oxidativa (eco00190)
3.6.1.27	Undecaprenil-difosfato fosfatase	<i>ybeQ</i> <i>bacA</i>	2	Núcleo	Biosíntese de peptidoglicanos (eco00550)
4.2.1.2	Fumarato hidratase	<i>fumA<sup>d</sup></i> <i>fumB<sup>d</sup></i> <i>fumC</i>	2	Núcleo	Ciclo do citrato (eco00020)*
4.1.1.19	Arginina decarboxilase	<i>cyaA</i> <i>ygiF</i>	2	Núcleo	Metabolismo de arginina e prolina (eco00330)
4.2.99.1 8	DNA liase (sítioapurínico ou apirimidínico)	<i>Nth</i> <i>Fpg<sup>d</sup></i> <i>NeI<sup>d</sup></i>	2	Núcleo	Reparo por excisão de bases (eco03410)
4.6.1.1	Adenilato ciclase	<i>cyaA</i> <i>ygiF</i>	2	Núcleo	Metabolismo de purinas (ec00230)
5.2.1.8	Peptidilproil isomerase	<i>ppiA<sup>e</sup></i> <i>ppiB<sup>e</sup></i> <i>fkIB</i> <i>slyD</i> <i>fkpA</i> <i>fkpB</i> <i>surA</i> <i>ppiD</i>	3	Núcleo	N/D
5.3.1.6	Ribose-5-fosfato isomerase	<i>rpiA</i> <i>rpiB</i>	2	Acessório	N/D
5.3.3.10	5-carboximetil-2-hidroxiuconato delta-isomerase	<i>hpaF</i> <i>ycgM</i>	2	Acessório	Metabolismo da tirosina (ec00350)
4.2.1	Fosfoglicerato mutase	<i>gpmA<sup>d</sup></i> <i>gpmB<sup>d</sup></i> <i>gpmI</i>	2	Núcleo	N/D

e: essencial, d: parálogo



Através da análise das variações intra-específicas das NISES nestes 33 genomas de *E. coli* (Tabela S3), observamos variações no padrão de presença e ausência de algumas formas enzimáticas presentes em uma determinada atividade enzimática. Como, neste trabalho, estamos investigando possíveis diferenças entre o *papel biológico atribuído* a uma determinada enzima, e cada atividade enzimática desempenha uma função metabólica distinta, examinamos em detalhes cada um dos casos de NISE encontrados, incluindo nesta análise os resultados de outros métodos analíticos empregados. A Tabela 3 resume os resultados das análises efetuadas para cada NISE identificada.

Tabela 3. Quadro comparativo dos métodos analíticos empregados neste trabalho.

PPA: padrão de presença/ausência, CG, Contexto genômico, EP: Estrutura proteína, PFQ: Padrão Físico e químico, PB: Padrão bioquímico, L: Localização, Rede de interação, FIT: Fitness, Essencialidade, EXP: expressão, CE: Co-expressão, R, Regulação da expressão, LM: Mineração da literatura, PB: Papel Biológico, ND: Não definido, D: diferente, N: Análise não realizada, Quadro verde, indica que os dados evidenciam distintos papéis biológicos; Quadro amarelo, análise não conclusiva; Quadro branco: análise não realizada; Quadro vermelho indica que não houve diferenças. .

EC	Atividade enzimática	CG	EP	PFQ	PB	L	RI	FIT	E	EXP	CE	R	LM	PB
1.1.1.1	Álcool desidrogenase	Red	Ver	Ver	Am	Ver	D							
1.1.1.28	D-lactato desidrogenase	Red	Ver	Ver	Am	Ver	D							
1.1.1.205	Inosine-5'-monophosphate dehydrogenase	Ver	Ver	Ver	Am	Ver	Ver	Ver	Ver	Am	Ver	Ver	Ver	D
1.1.1.52	Quinoprotein glucose desidrogenase (PQQ, quinone)	Red	Ver	Ver	Am	Ver	Ver	Ver	Ver	Am	Ver	Ver	Ver	D
1.97.1.9	Selenate redutase	Ver	Ver	Ver	Am	Ver	Ver	Ver	Ver	Am	Ver	Ver	Ver	ND
1.11.1.6	Catalase	Red	Ver	D										
1.15.1.1	Superoxido dismutase	Ver	Ver	Ver	Am	Ver	D							
1.5.1.29*	FMN redutase (NADPH)	Ver	Ver	Ver	Am	Ver	Ver	Ver	Ver	Am	Ver	Ver	Ver	D
1.5.1.3	Di-hidrofolate reductase	Ver	Ver	Ver	Am	Ver	Am	D						
2.3.1.18	Galactosideo O-acetiltransferase	Red	Ver	Ver	Am	Ver	Ver	Ver	Ver	Am	Ver	Ver	Ver	ND
2.1.2.2	Phosphoribosilglicinamida formiltransferase	Red	Ver	Ver	Am	Ver	D							
2.3.1.15	Glicerol-3-fosphate 1-O-aciltransferase	Ver	Ver	Ver	Am	Ver	D							
2.5.1.17	Ácido Cobirínico a,c-diamida adenosiltransferase	Ver	Ver	Ver	Am	Ver	Ver	Ver	Ver	Am	Ver	Ver	Ver	D
2.7.1.11	6-fosfofrutocinase isozime	Red	Ver	Ver	Am	Ver	D							
2.7.1.4	Fructocinase.	Ver	Ver	Ver	Am	Ver	ND							
3.1.1.5	Lisofosfolipase	Red	Ver	Ver	Am	Ver	Ver	Ver	Ver	Am	Ver	Ver	Ver	D
3.1.22.4	<i>Crossover junction endodeoxyribonuclease</i>	Ver	Ver	Ver	Am	Ver	D							
3.1.3.1	Alcaline fosfatase	Red	Ver	Ver	Am	Ver	Ver	Ver	Ver	Am	Ver	Ver	Ver	D
3.1.3.2	Fosfatase ácida	Red	Ver	Ver	Am	Ver	D							
3.1.3.27	Fosfatidilglicerofosfatase	Red	Ver	Ver	Am	Ver	D							
3.1.3.5	5'-nucleotidase	Ver	Ver	Ver	Am	Ver	D							
3.5.2.6	Beta-lactamase	Red	Ver	Ver	Am	Ver	Ver	Ver	Ver	Am	Ver	Ver	Ver	ND
3.6.1.1	Inorgânica difosfatase	Red	Ver	Ver	Am	Ver	D							
3.2.1.17	Lisozima	Ver	Ver	Ver	Am	Ver	Ver	Ver	Ver	Am	Ver	Ver	Ver	D
3.2.1.22	Alfa-galactosidase.	Ver	Ver	Ver	Am	Ver	ND							
3.2.1.86	6-fosfo-beta-glucosidase	Ver	Ver	Ver	Am	Ver	D							
3.6.1.27	<i>Undecaprenil-dijsfato fosfatase</i>	Red	Ver	Ver	Am	Ver	ND							
4.2.1.2	Fumarato hidratase	Red	Ver	Ver	Am	Ver	Ver	Ver	Ver	Am	Ver	Ver	Ver	D
4.1.1.19	Arginina decarboxilase	Red	Ver	Ver	Am	Ver	D							
4.2.99.18	<i>DNA-(apurinic or apyrimidinic site) lyase</i>	Red	Ver	Ver	Am	Ver	Ver	Ver	Ver	Am	Ver	Ver	Ver	ND
4.6.1.1	Adenilate cyclase	Red	Ver	Ver	Am	Ver	Ver	Ver	Ver	Am	Ver	Ver	Ver	ND
5.2.1.8	Peptidilprolil isomerase.	Ver	Ver	Ver	Am	Ver	D							
5.3.1.6	Ribose-5-Phosphate isomerase	Red	Ver	Ver	Am	Ver	D							
5.3.3.10	5-carboximetil-2-hidroxiuconate delta-isomerase.	Ver	Ver	Ver	Am	Ver	ND							
5.4.2.1	Fosfoglicerato mutase	Ver	Ver	Ver	Am	Ver	ND							

A análise do padrão de presença e ausência permitiu identificar as enzimas do genoma núcleo e do genoma acessório, e a análise do padrão de distribuição de formas estruturais (enovelamentos) revelou um aumento do número de cópias de algumas enzimas em linhagens patogênicas de *E. coli* (Tabela S3).

A análise da taxa de crescimento (*fitness*) dos mutantes (Tabela S4) e dados de essencialidade (Tabela S5) compilados na literatura permitiram avaliar a importância de cada gene para a sobrevivência do organismo. Por exemplo, para a atividade enzimática Glicerol-3-fosfato 1-O-aciltransferase foram identificados três análogos funcionais, mas apenas o gene que codifica *plsB* é essencial para *E. coli* (Tabela S5).

Os resultados das redes de interação de análogos intragenômicos foram baseados nas múltiplas fontes de evidências experimentais que estão disponíveis no banco de dados *STRING*, permitindo uma avaliação de sua função geral no metabolismo.

Os resultados das análises dos padrões de expressão (Tabela S6) e co-expressão gênica (Tabela S7) em *E. coli* sob diferentes condições fisiológicas tais como: anaerobiose/aerobiose, anaerobiose/variações de pH, fase de crescimento exponencial/fase de crescimento estacionário e formação de biofilme indicam que o papel biológico das NISEs intragenômicas pode ser distinto e variar de acordo com as condições fisiológicas. Em determinada condição fisiológica um gene que codifica uma forma estrutural pode ser expresso enquanto o gene análogo pode ser inibido. Por exemplo, os análogos *adhP* e *adhE*, *gpmM* e *gpmA*, *sodC* e *sodA* possuem distintos padrões de expressão em *E. coli* durante condições de crescimento anaeróbico (Tabela S5).

Avaliamos alguns parâmetros físicos e químicos de NISEs através do peso molecular, número total de átomos, ponto isoelétrico teórico (PI), composição de cada aminoácido (%), total de resíduos carregados positivamente ou negativamente, índice de estabilidade, índice alifático e grande média de hidropaticidade (GRAVY) (Tabela S9)

Além das diferenças nos parâmetros físicos e químicos, algumas formas estruturais estão localizadas em distintos compartimentos celulares (Tabela S8 e S9). As proteínas localizadas na membrana plasmática requerem menos gasto de energia para sua síntese independentemente de seu comprimento, função e estrutura, pois os aminoácidos utilizados geralmente são menos custosos (Smith e Chapman, 2010).

Algumas formas análogas possuem distintos parâmetros bioquímicos, tais como: cofatores, substratos, valores de constante de Michaelis ( $K_m$ , onde  $< K_m = >$  afinidade), localização celular e pH ótimo para atividade enzimática (Tabela S8). Algumas isoenzimas que têm evoluído sobre condições metabólicas ligeiramente diferentes sugerem que esta variação pode ser adaptativa, como por exemplo, as isoenzimas lactato desidrogenase e glicerol 3-fosfato aciltransferase, que são encontradas em tecidos do fígado e do músculo; os valores de  $K_m$  encontrado para as enzimas do músculo são duas vezes mais altos do que para as enzimas do fígado, o que é consistente com valores de  $K_m$  adaptados para altas concentrações fisiológicas dos dois substratos no músculo (Benner, 1989). Existem vários argumentos persuasivos que mostram que a seleção natural pode ajustar o valor de  $K_m$  e isto pode resultar em adaptação. Enzimas com baixos valores de  $K_m$  podem ser positivamente selecionadas através da seleção natural, para mais detalhes ver a revisão de Benner (1989).

Os sistemas metabólicos geralmente preferem expressar genes que são menos custosos e a seleção natural parece favorecer as proteínas curtas (Akashi & Gojobori, 2002). As proteínas menores geralmente são mais expressas, devido ao menor custo metabólico para sua síntese (Akashi, 2003), as proteínas com composição similares de aminoácidos tendem a ter igual padrão de expressão (Raghava & Han, 2005). De fato, observamos que tamanho das algumas NISEs é o custo para sua síntese é distinto, como por exemplo, os pares de análogos AppA e AphA (Fosfatase Ácida), SurE e UshA (5'-nucleotidase), ChiA e YafL (Lisozima), GpmA e GpmM (Fosfoglicerato mutase) (Tabela S9).

A comparação de alguns modelos tridimensionais de algumas enzimas isofuncionais não homólogas (NISES) ou análogas em *E. coli* confirmaram os distintos padrões de enovelamentos adotados por estas proteínas (Figuras S1).

## 4.2 Superóxido dismutase (EC 1.15.1.1): SodA, SodB, SodC

### 4.2.1 Número de cópias e contexto genômico

Através de análises do padrão de presença e ausência de formas estruturais (enovelamentos) enzimáticas com casos de analogia em *E. coli*, identificamos um aumento no número de cópias do gene *sodC* em três (de quatro) linhagens enterohemorrágicas (EHEC): O157: H7 EDL933, O157: H7 Sakai, O103: H2 12009 e O111:H-11128 (Figura 2). Os genes *sodA* e *sodB* codificam enzimas Superóxido dismutases que possuem um mesmo padrão de enovelamento estrutural (Tabela S2), mas que possuem distintos requerimentos quanto aos cofatores utilizados (Tabela S8).

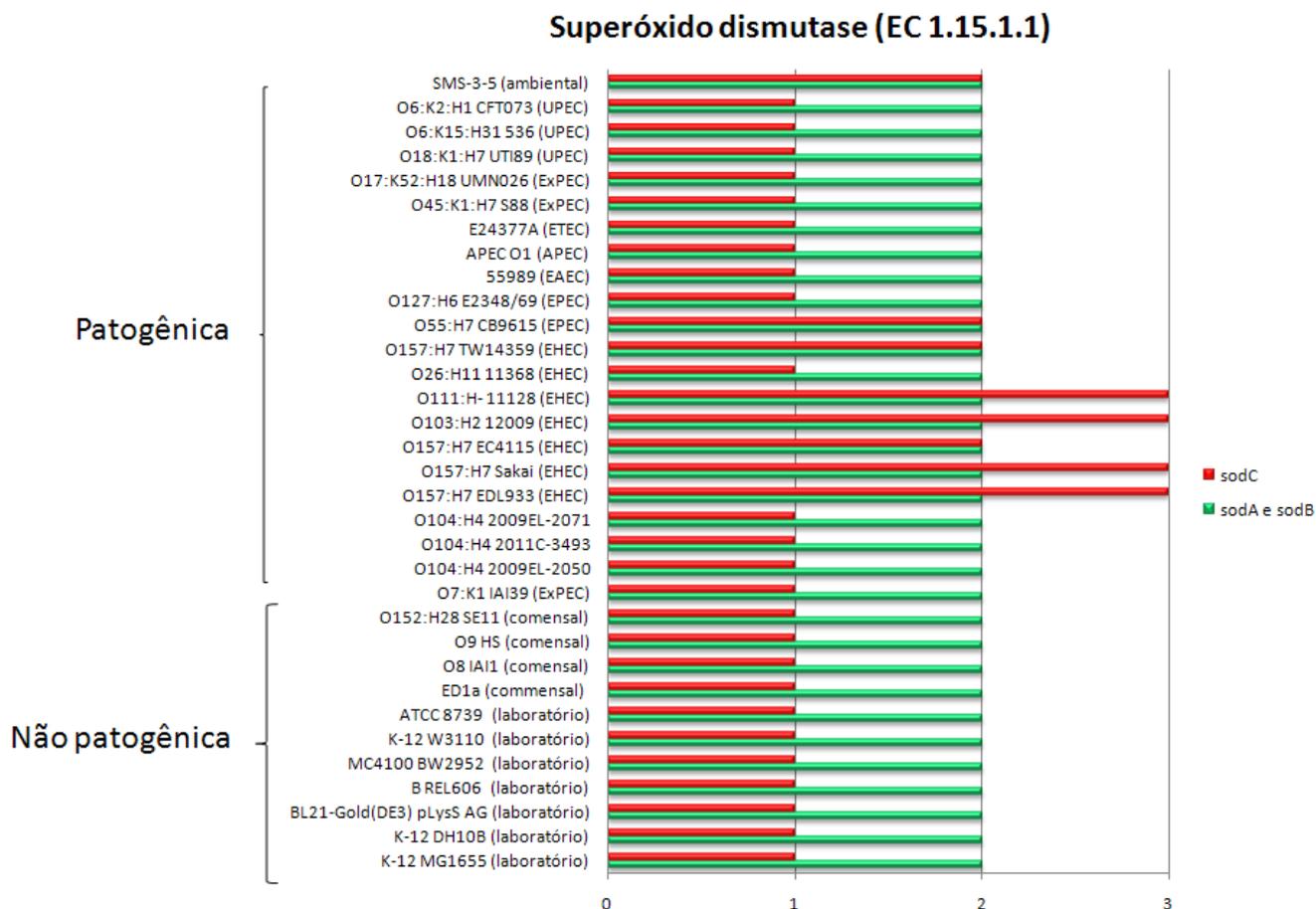
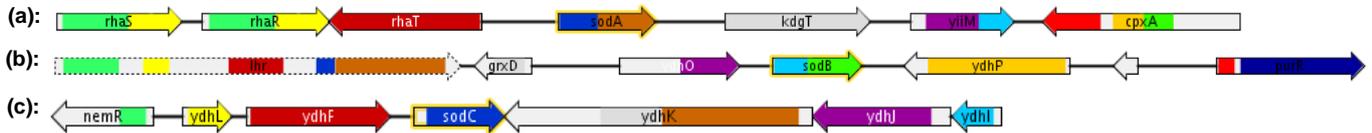


Figura 2 – Número de cópias de *sodA*, *sodB* e *sodC*. Barras vermelhas: *sodC*; barras verdes: *sodA* e *sodB*.

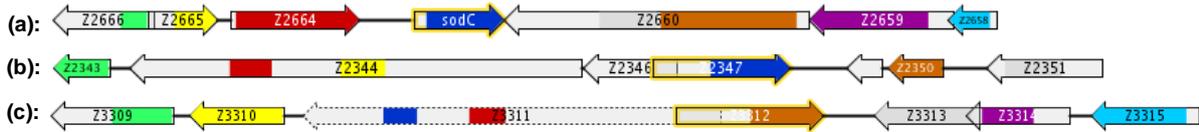
Na linhagem *E. coli* K12 o gene *sodA* está localizado na posição 4,098,833 -> 4,099,453 do genoma, *sodB* está localizado na posição 1. 733.402 -> 1. 733.983 e *sodC* está localizado na posição 1.722.158 <- 1.722.679. A linhagem *E. coli* K12 não patogênica possui apenas uma cópia dos genes *sodA*, *sodB* e *sodC* (Figura 3A). Identificamos duas cópias de *sodC*, nas seguintes linhagens: SMS-3-5 (ambiental) (Figura 3D), O157: H7 EC4115 (EHEC), O157: H7 TW14359 (EHEC) e O55: H7 CB9615 (EPEC) e três cópias de *sodC* nas linhagens enterohemorrágicas: O157: H7 EDL933 (EHEC) (Figura 3B), O157: H7 Sakai (EHEC) (Figura 3C), O103: H2 12009 (EHEC) e O111: H-11128 (EHEC).

Na linhagem O157: H7 EDL933, duas cópias de *sodC* são encontradas no profago CP-933R; na linhagem O157: H7 EDL933, duas cópias são encontradas dentro de um profago não anotado no bando de dados GENE (NCBI) e na linhagem SMS-3-5, *upstream* ao gene *sodC*, identificamos uma sequência que codifica uma proteína de uma sequência de inserção (lócus EcSMS35\_3164), seguida por uma sequência que codifica uma transposase (lócus: EcSMS35\_3165); a completa anotação do contexto genômico é descrita nos anexos (Figuras S2).

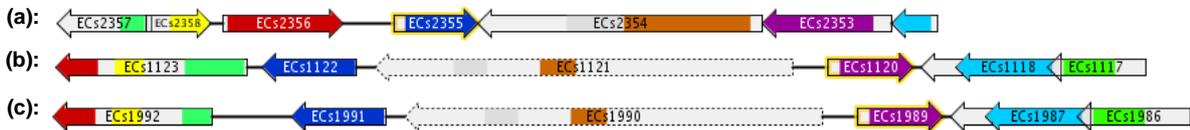
**A: Contexto genômico local dos genes sodA (a), sodB (b) e sodC (c) em *E. coli* K12**



**B: Contexto genômico sodC em *E. coli* O157:H7 EDL933 (EHEC)**



**C: Contexto genômico sodC em *E. coli* O157: H7 Sakai (EHEC)**



**D: Contexto genômico sodC em *E. coli* SMS-3-5 (environmental)**

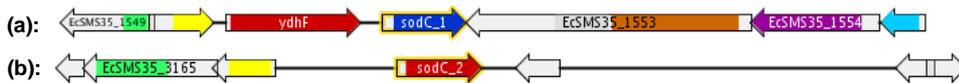


Figura 3 – Contexto genômico local de *sodC* nas Linhagens K12 (Fig. 3A), O157: H7 EDL933 (EHEC) (Fig. 3B), O157: H7 Sakai (EHEC) (Figura 3C) e SMS-3-5 (ambiental) (Fig. 3D). Na Fig. 3 A, na letra (a) *sodA* é representado com contorno amarelo, na Fig. 3A em (b) *sodB* é representado com contorno amarelo, na Fig. 3A em (c) *sodC* é representado com contorno amarelo. A Figura 3B, letra (a), (b) e (c) mostra as três cópias de *sodC* com contorno amarelo em diferentes locus do genoma de O157: H7 EDL933 (EHEC). A Fig 3C, letra (a), (b) e (c) mostra as três cópias de *sodC* com contorno amarelo em diferentes locus do genoma de O157: H7 Sakai (EHEC) e a Fig 3D mostra as três cópias de *sodC* com contorno amarelo em diferentes locus do genoma de SMS-3-5 (ambiental).

#### 4.2.2- Análise do perfil de expressão gênica de *sodA*, *sodB* e *sodC*

Com o *programa* Genevestigator (Hruz et al., 2008) analisamos o perfil de expressão dos genes *sodA*, *sodB* e *sodC* em diferentes condições fisiológicas, a partir de dados disponíveis no banco de dados *Gene Expression Omnibus* ([www.ncbi.nlm.nih.gov/geo/](http://www.ncbi.nlm.nih.gov/geo/)). Estes mesmos dados foram compilados e também re-analisados com o *programa* R.

A análise dos dados de expressão gênica revelou que *sodA* (P00448) é reprimido durante condições de anaerobiose e induzido em condições de aerobiose K12 (Figura 4), *sodB* (P0AGD2) é induzido durante condições de anaerobiose (Figura 5). Em condições anaeróbicas *sodB* é única enzima presente em *E. coli* (Kargalioglu & Imlay, 1994).

Em condições de anaerobiose e pH ácido (5.7) e pH básico (8.5), *sodC* passa a ser induzido, assim como *sodB* (em um nível menor), ou seja, na presença de pH ácido a expressão de *sodC* é induzida mesmo sob condições de anaerobiose, isto indica que pequenas alterações no ambiente podem possivelmente alterar o papel biológico de um gene. Em *Y. enterocolitica* as enzimas SodA e SodB possuem pH ótimo de atividade baixo, isto implica seu papel em ambientes ácidos tais como vesículas intrafagocíticas (Dhar, Gupta, & Viridi, 2013). Na linhagem *E. coli* uropatogênica UT189 o gene *sodC* também é reprimido durante condições de anaerobiose (Figura 5). Em condições de fase estacionária *sodC* é induzido e *sodB* é reprimido.

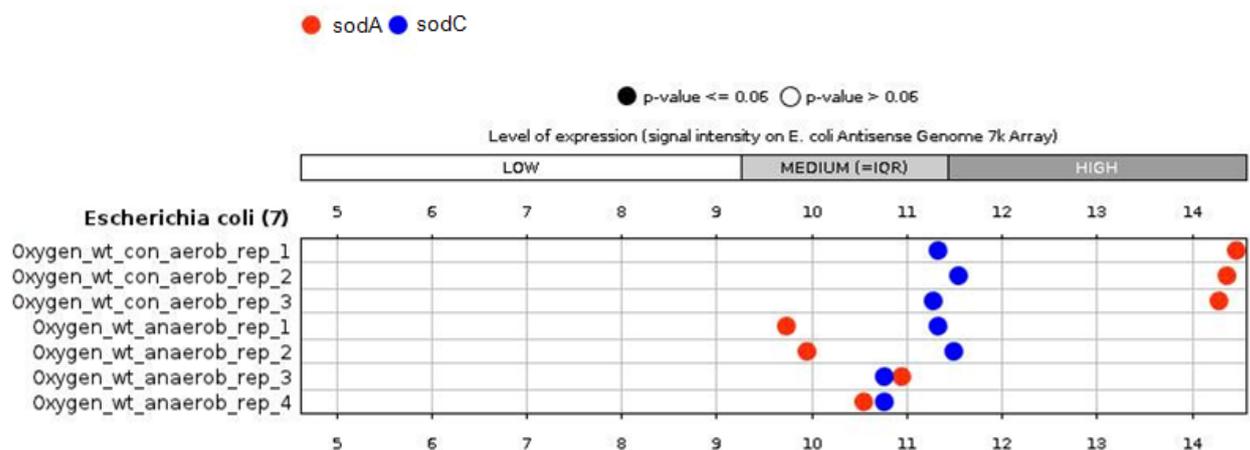


Figura 4 – Padrão de expressão de *sodA* e *sodC*.

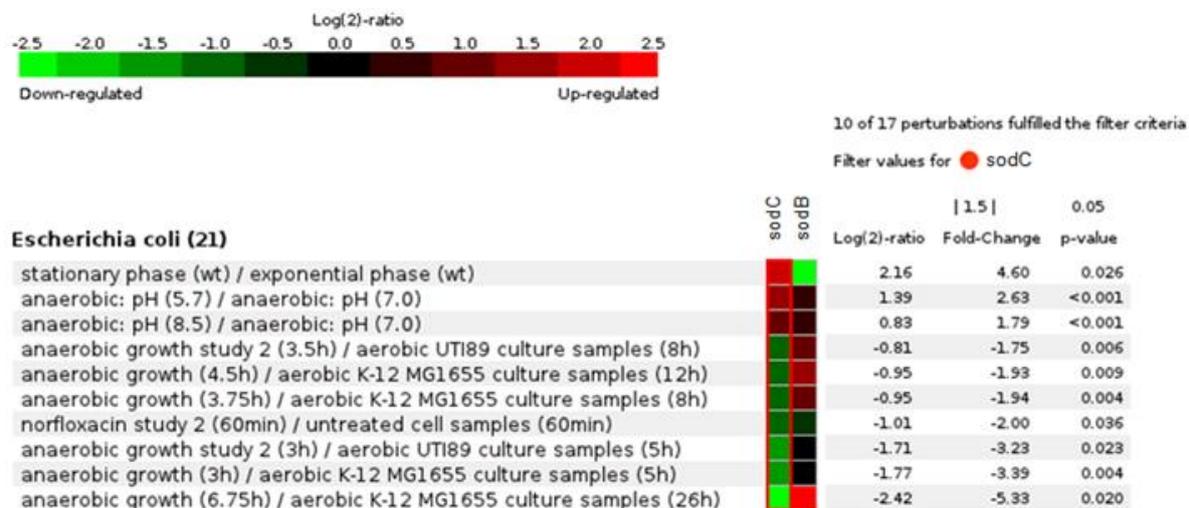


Figura 5 – Padrão de expressão de *sodB* e *sodC*.

#### 4.2.3 Análise dos dados e literatura

Os microorganismos aeróbicos possuem enzimas antioxidantes, tais como superóxido dismutases e catalases. Estas enzimas propiciam a proteção dos danos diretos e indiretos causados por espécies reativas de oxigênio (ROS) tais como  $O_2^-$  e  $H_2O_2$ . Estudos mostram que *E. coli* mutante para superóxido dismutases são mais suscetíveis a efeitos deletérios causados por  $H_2O_2$  no DNA, proteínas e lipídios (McCormick, Buettner, Britigan, & Cormick, 1998). Em *Salmonella* spp., *sodC* desempenha papel chave na patogenicidade através da interceptação de espécies reativas de oxigênio produzidas pela resposta imune do hospedeiro (Uzzau, Bossi, & Figueroa-Bossi, 2002).

Em *Mycobacterium tuberculosis*, SodC protege a membrana da bactéria de espécies reativas de oxigênio (ERO) geradas no ambiente externo; bactérias com este gene deletado não sobrevivem à resposta oxidativa produzida por macrófagos (Piddington *et al.*, 2001). A presença de um análogo funcional no genoma de *E. coli* pode beneficiar o microorganismo contra os efeitos deletérios de espécies reativas de oxigênio.

A enzima SodC está relacionada com virulência em *Salmonella typhimurium* (Korshunov & Imlay, 2002), *Cryptococcus neoformans* (G. M. Cox et al., 2003) e outras bactérias. Alguns estudos indicam que SodC possui um papel de proteção em bactérias durante sua infecção na célula hospedeira, pois a enzima atua contra espécies reativas de oxigênio geradas por macrófagos; isto foi corroborado por análises de mutantes deficientes para este gene, pois estas são menos virulentas que a bactéria selvagem (Battistoni, 2003). Em *Burkholderia pseudomallei*, a deleção de *sodC* diminui a sobrevivência da bactéria em células de macrófagos de camundongo (Vanaporn et al., 2011). O gene *sodC* já foi isolado de *Salmonella* sp. resistentes à drogas em pacientes sofrendo de febre entérica (Sanjay et al., 2010)

Estes dados sugerem que o papel biológico de *sodC* em bactérias é auxiliar na defesa e proteção contra a resposta imune do hospedeiro, em contraposição ao seu análogo, *sodB*, envolvido na remoção de radicais livres e superóxidos resultantes do próprio metabolismo. Esse papel protetor possivelmente ocorre devido à sua localização no periplasma da célula bacteriana (Piddington et al., 2001). Na linhagem ambiental de *E. coli* SMS-3-5, *sodC* pode possivelmente desempenhar um papel biológico contra espécies reativas de oxigênio produzidas em condições tóxicas. Esta linhagem foi isolada a partir de um estaleiro na costa de Creek, Charleston, Carolina do Sul, um ambiente contaminado com metais industriais tóxicos. É resistente a um número recorde de antibióticos, em muitos casos, em altas concentrações ([http://gsc.jcvi.org/projects/msc/e\\_coli\\_and\\_shigella/](http://gsc.jcvi.org/projects/msc/e_coli_and_shigella/)).

Os dados em relação ao contexto genômico das distintas formas alternativas da superóxido dismutase sugerem que o aumento do número de cópias de *sodC* nas linhagens patogênicas de *E. coli* pode estar relacionado à integração de profagos no genoma; neste cenário, os profagos carregam uma cópia adicional de *sodC* para estas bactérias. Em um cenário alternativo, devido à presença proximal de transposases, eventos de transferência horizontal seguidos por transposição no genoma também podem estar associados ao aumento do número de cópias de *sodC*. De qualquer forma, tanto o número de cópias quanto sua distribuição são aparentemente distintos entre *sodC* e *sodA*, comportamento este verificado em distintas linhagens de *E. coli*.

Os dados de expressão gênica obtidos no *programa* Genevestigator indicam que tanto *sodC* e *sodA* são similares em alguns aspectos: ambas, por exemplo, são reprimidas pelo antibiótico norflaxacin (Figura 5). O pH também influencia na indução de expressão dos genes que codificam para estas enzimas, em condições de crescimento anaeróbico: nestes experimentos, os dados mostraram diferenças significativas entre seus padrões de expressão. Em determinadas condições, a expressão ou presença de uma forma enzimática específica no citoplasma pode fornecer maior plasticidade e adaptabilidade à condição imposta. Neste contexto, os resultados de expressão e co-expressão mostram que as diferentes superóxidos dismutases podem ser recrutadas em momentos ou condições fisiológicas distintas por *E. coli*, indicando que seu papel biológico é claramente distinto.

Como esperado, *sodA* e *sodC*, as redes de interação obtidas (dados nos anexos) mostram que ambas interagem com distintas proteínas. De acordo com a literatura, estas proteínas possuem convergência estrutural em resíduos do sítio ativo e também mecanismos similares de transformação química (Gherardini, Wass, Helmer-Citterich, & Sternberg, 2007a), mas por outro lado o distinto enovelamento da estrutura secundária entre ambas as formas, sugere uma distinta origem evolucionária (Tabela S2).

#### **4.3 Dihidrofolato redutase (EC 1.5.1.3): FoIA e FoIM**

A dihidrofolato redutase (EC 1.5.1.3) atua na síntese e manutenção dos níveis de tetrahydrofolato, uma molécula que é essencial para síntese de aminoácidos, purinas e timidilato e conseqüentemente para o crescimento celular (Schnell et al., 2004). Nós identificamos duas cópias da forma enzimática FoIA (Figura 6) nos genomas de linhagens causadoras de surtos endêmicos na Georgia e Alemanha. Estas linhagens causaram recentes surtos de diarreia hemorrágica em humanos, com vários casos fatais (<http://www.cdc.gov/ecoli/general/germany.html>); tais bactérias possuem resistência contra vários antibióticos da classe cefalosporina, o que está correlacionado com a presença de uma cópia da enzima cefalosporinase no plasmídeo Inc1 (Ahmed *et al.*, 2012).

### Dihidrofolato redutase (EC 1.5.1.3)

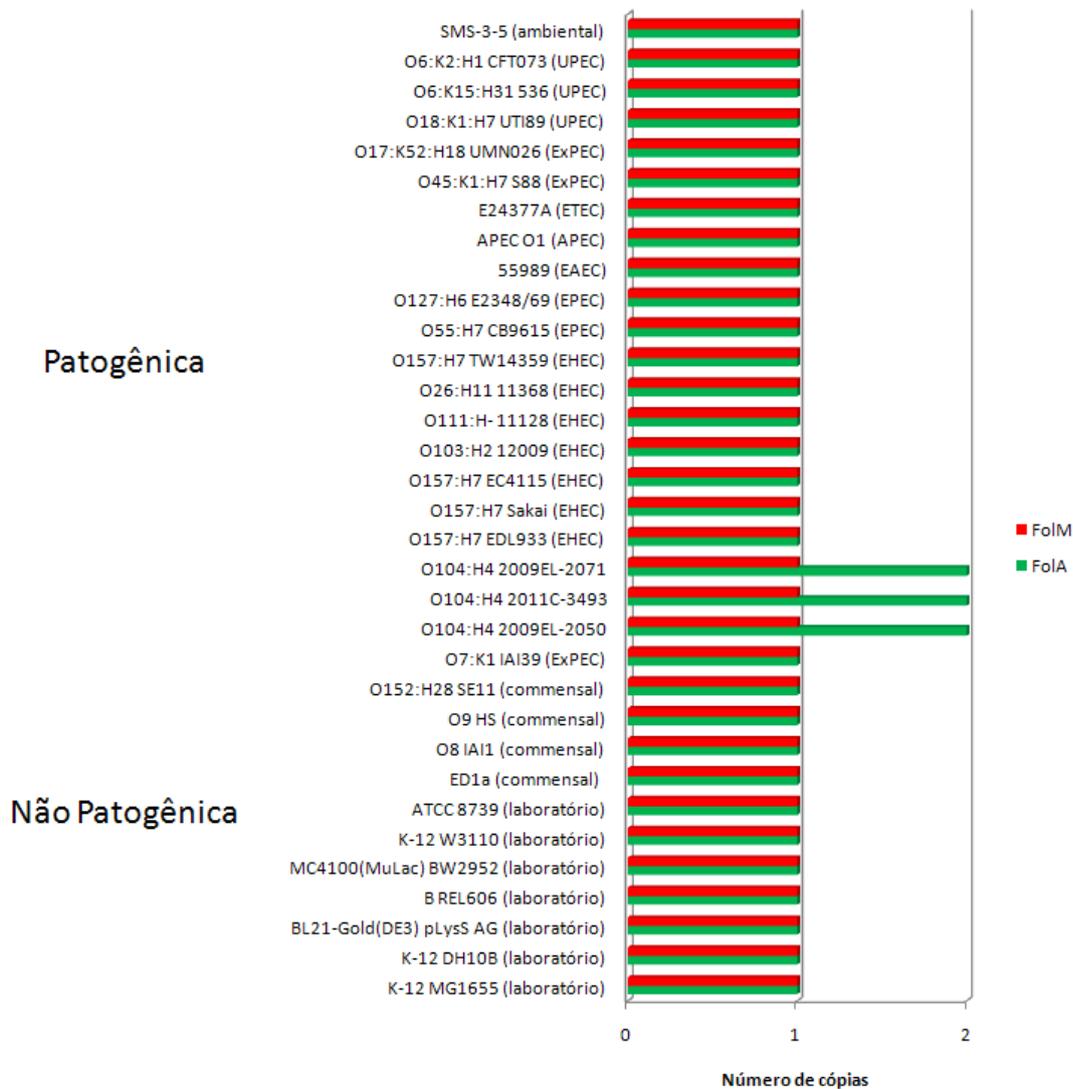


Figura 6 – Número de cópias de *foIA* e *foIM* nas 33 linhagens de *E. coli* estudadas. EHEC: Enterohemorrágica, EPEC: Enteropatogênica, ExPEC: Estraintestinal, ETEC: Enterotoxigênica.

A cópia adicional de *foIA* está localizada em uma distinta região do genoma onde vários genes vizinhos codificam transposases e também uma integrase, o que sugere que este gene pode ter surgido a partir de um evento de transposição no genoma (Figura 7). O gene *foIM* não está amplificado em *E. coli* e seu contexto genômico é conservado entre linhagens patogênicas e não patogênicas (Figura 8).

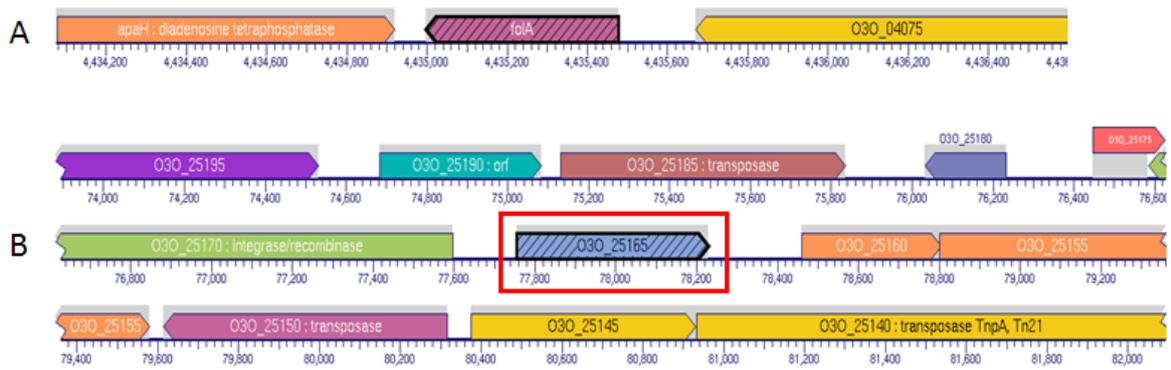


Figura 7 – Contexto genômico das duas cópias de *folA* em *E. coli* O104: H4 2009 EL-2050: A: Cópia 1 de *folA*. B: Cópia 2 de *folA* (em destaque): Genes *upstream* ao gene *folA*1: transposase (O30\_25185), integrase/recombinase (O30\_25170). Genes *downstream* ao gene *folA*2: transposase (O30\_25150), **Transposase TnpA, Tn21** (O30\_25140).

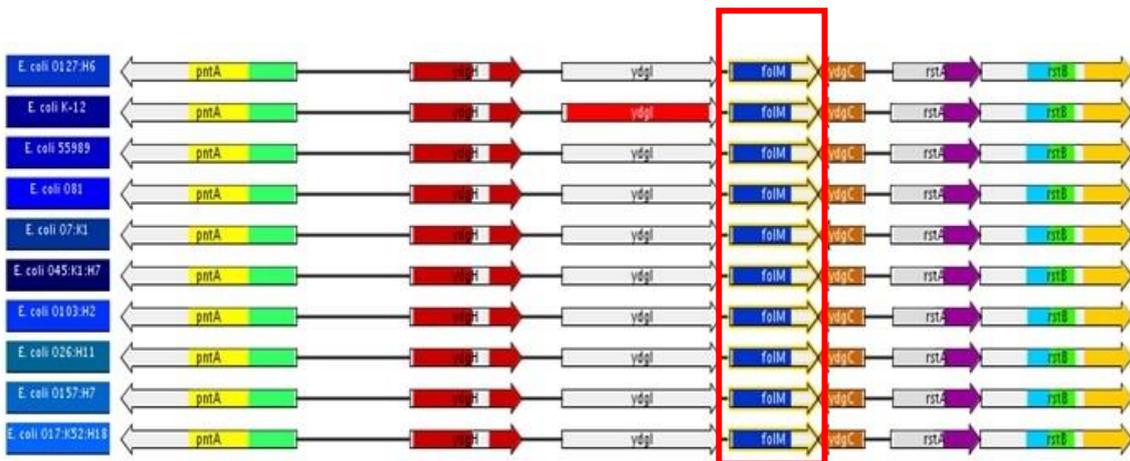


Figura 8 – Contexto genômico de *folM* em algumas linhagens patogênicas e não patogênicas de *E. coli*. A descrição completa dos genes encontra-se na *FiguraS2*

#### 4.3.1 Análise da taxa de sobrevivência de mutantes (*E. coli*) *folA* e *folM* quando submetidas à diferentes estresses ou compostos químicos

Através do banco de dados PortEco (<http://www.porteco.org/>) obtivemos os dados do fenótipo de mutantes para o gene *foM* em *E. coli* BW25113 (Keio Collection, <http://ecoli.aist-nara.ac.jp/gb6/Resources/deletion/deletion.html>) e analisamos os dados de sobrevivência e adaptação (*Fitness*) em várias condições de estresses e compostos químicos. A atividade enzimática mediada por FoIA pode ser inibida pelos antibióticos *trimetropim*, metotrexato e *Sulfamethizole*, enquanto FoIM é inibida apenas por metotrexato (Giladi *et al.*, 2003). Entretanto, os resultados da taxa de *fitness* evidenciam um crescimento positivo no mutante *foM* mesmo na presença de trimetropim (Figura 9).

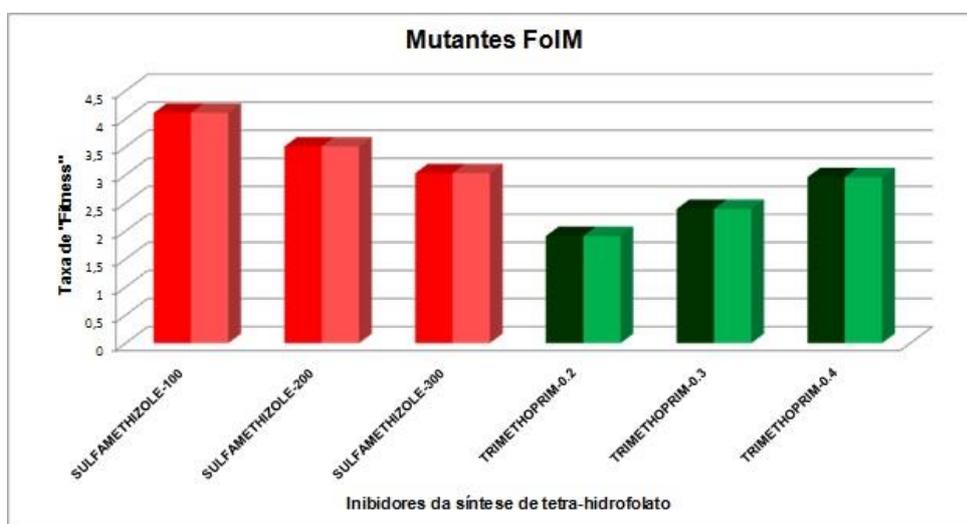


Figura 9 – Taxa de *fitness* do mutante *foM* de *E. coli* BW25113 tratados com inibidores da síntese de tetra-hidrofolato. As linhas com as mesmas cores em diferentes intensidades representam réplicas, tratadas com concentrações crescentes dos antibióticos *Sulfamethizole* e *Trimetoprim*. Dados obtidos do banco de dados PortEco (<http://www.porteco.org/>).

A análise da rede de interação de FoIA em *E. coli* K12 indica que FoIA é uma enzima central (maior grau de conectividade) e que possui maior influência (maior valor de *betweenness*) no metabolismo de dihidrofolato (Figura 10). Para entender melhor o papel biológico de FoIM, obtivemos a rede de interação desta proteína através do banco dos dados *STRING* e calculamos o grau de interação e *betweenness* para todas as proteínas da rede. Para esta proteína nós excluímos o análogo funcional FoIA para mapear apenas as interações envolvendo FoIM. A

análise da rede biológica de FolM mostra que esta proteína interage com enzimas envolvidas no metabolismo de purinas e pirimidinas (Figura 11).

Com relação às proteínas que participam da rede biológica de folA, as enzimas GcvP, GcvT, GlyA e MetH estão envolvidas no metabolismo de aminoácidos; FolC, PurN, PurT e PurU fazem parte do metabolismo de tetra-hidrofolato; FolX é requerida para síntese de tetra-hidromonapterina, ThyA catalisa a conversão de dUMP para dTMP; LpdA é um componente de três multicomplexos enzimáticos: a piruvato desidrogenase, o complexo 2-oxoglutarato desidrogenase e o sistema de clivagem de glicina (Steiert *et al.*,1990).

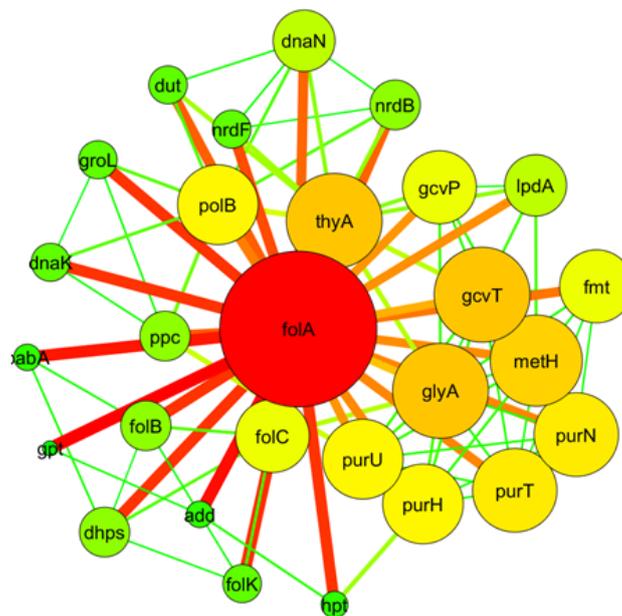


Figura 10 – Rede de interação de FolA. (A) Rede de interação de FolA em *E. coli* K12 obtida no banco de dados *STRING* (*STRING* -db.org/) e geradas através de análises no programa *Network analysis*, implementado no programa *Cytoscape* (versão 2.8), o tamanho dos nós e intensidade das cores correspondem a intensidade dos valores obtidos das medidas grau de interação e *betweenness*.

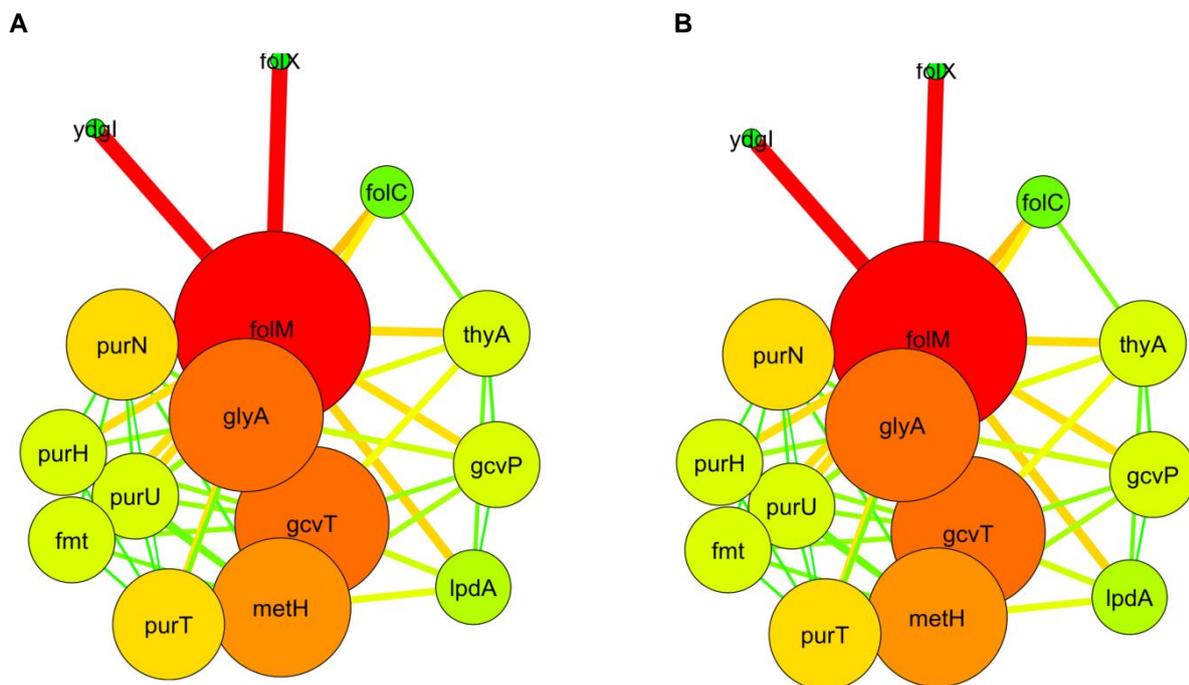


Figura 11 – Rede de interação de FoIM. (A) Rede de interação de FoIM em *E. coli* K12 obtida no banco de dados *STRING* (*STRING* -db.org/) e geradas através de análises no programa *Network analysis*, implementado no programa *Cytoscape* (versão 2.8), o tamanho dos nós e intensidade das cores correspondem a intensidade dos valores obtidos das medidas grau de interação e *betweenness*.

#### 4.3.2 Parâmetros físico-químicos e comparação estrutural entre FoIA e FoIM

A enzima FoIM é uma proteína multifuncional que catalisa a redução de H<sub>2</sub>-folato (Di-hidrofolato) para H<sub>4</sub>-folato (Tetrahidrofolato) (Giladi *et al.*, 2003) e de H<sub>2</sub>-MP (Dihidromonapterina) para H<sub>4</sub>-MP (tetra-hidromonapterina), um cofator para a enzima PhhA (fenilalanina hidroxilase). FoIA atua especificamente na redução de dihidrofolato e possui alta afinidade por este substrato, maior K<sub>m</sub> (Michaelis constant) que sua forma análoga FoIM (Tabela S8). A transformação química mediada por FoIA é essencial para a manutenção das concentrações intracelulares de Di-hidrofolato.

O modelo 3D de FoIA foi obtido do banco de dados PDB e o de sua forma análoga FoIM através de modelagem comparativa com a sequência (3GEM)

homóloga de *Pseudomonas syringae* pv. *phaseolicola* 1448 (Figura 12). A modelagem comparativa confirma que tais enzimas são análogas, pois apresentam distintos enovelamentos em suas estruturas terciárias. A enzima FoIA é um homodímero que possui o resíduo Asp27 no sítio ativo, local de ligação do inibidor metotrexato (B. Bennett et al., 2006), outros resíduos importantes para a atividade catalítica são Leu5, Ile5, Met20, Phe31, Leu28, Ile94 (*Catalytic Site Atlas*, <http://www.ebi.ac.uk/thornton-srv/databases/CSA/>). A proteína FoIM é um membro da família SDR, nesta família os resíduos importantes são Gly14, Gly19, Pro181, Ala180, HIS189, TYR194, GLU188 (Jomvall et al., 1995). Os prováveis resíduos catalíticos de foIM são Arg18, Ser87, Trp89, Asp139, Thr152 e Lys156 (Giladi et al., 2003). Os dados da literatura sugerem que *foIA* e *foIM* possuem distintos resíduos no sítio ativo. Este é um caso particular onde análogos transformacionais possuem distintos resíduos no sítio ativo, pois, segundo Gherardini e colaboradores (2007), análogos transformacionais possuem resíduos idênticos ou similares (Gherardini et al., 2007). Os análogos transformacionais são enzimas que catalisam a mesma transformação química (convergência funcional) e possuem resíduos idênticos ou similares no sítio ativo (convergência mecânica) (Gherardini et al., 2007).

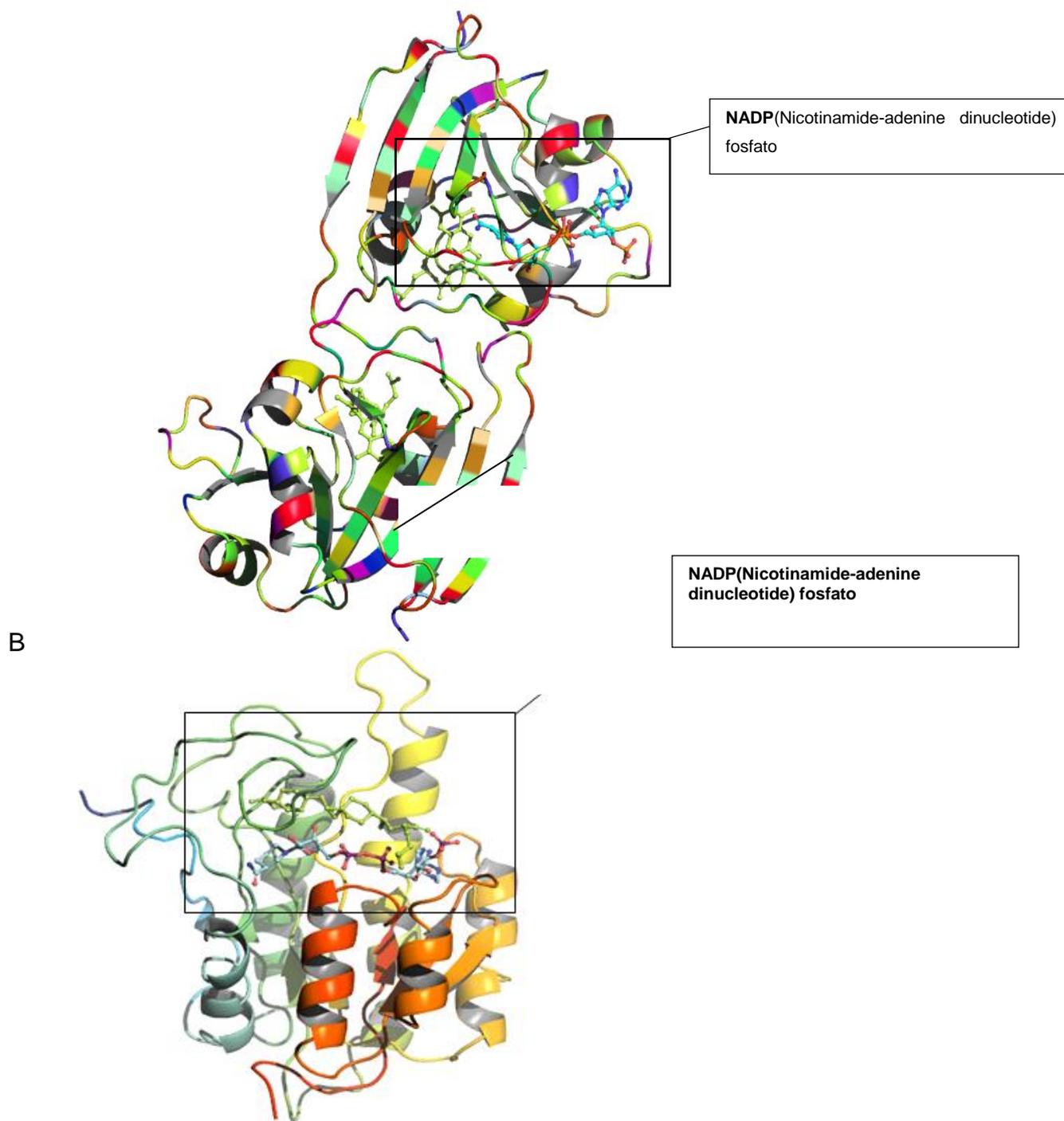


Figura 12 – Estrutura 3D de FolA e FolM mostrando distintos enovelamentos estruturais entre formas análogas de Dihidrolato redutase em *E. coli*. A: O homo-dímero FolA (P0ABQ4). (B): Modelo de FolM (P0AFS3) obtido na presença de metotrexato e do cofator NADP, os prováveis resíduos do sítio ativo são Arg18, Ser87, Trp89, Asp139, Thr152 e Lys156 em FolM (Giladi *et al.*, 2003). PDB [ID]: 1DDR, 3GEM molde para FolA e FolM, respectivamente. Ambas as figuras foram geradas pela interface do programa MAESTRO.

### 4.3.3 Análise dos dados e literatura

Em relação à estrutura da proteína FolM, as análises no banco de dados SUPERFAMILY revelaram que FolM e as sequências homólogas da classe oxidorreductase (PDB: 3GEM) de *Pseudomonas syringae* pv. *phaseolicola* 1448A e dihidropteridina redutase (DHPR) de *Homo sapiens* compartilham o mesmo enovelamento (NAD(P)-binding Rossmann-fold domains). Estas proteínas são membros da família *Short-Chain Dehydrogenases Reductases* (SDR), a qual tem como característica a conservação dos resíduos Tyr152 (ou Tyr151) e Lys 156, os quais são funcionalmente importantes (Jomvall *et al.*, 1995, Gourley *et al.*, 2001). Além destes, um motivo GlyXXXGlyXGly é característico das desidrogenases em geral formando um sítio de ligação à coenzimas (Rossmann *et al.*, 1975; Wierenga *et al.*, 1985). Um resíduo de ácido aspártico (Asp) regula a especificidade para NAD ou NADP<sup>+</sup> (Scrutton *et al.*, 1990). A dihidropteridina redutase (DHPR) de *H. sapiens* catalisa a transformação química da dihidrobiopterina para tetradihidrobiopterina, que é um cofator para hidroxilases de aminoácidos (Varughese *et al.*, 1992).

Análises bioquímicas mostram que a enzima multidomínio e multifuncional PTR1 (dihidrofolato redutase - timidilato sintase bifuncional) de *Leishmania major* catalisa a reação de redução NADPH-dependente de folato para tetrahydrofolato (Nare *et al.*, 1997, Wang *et al.*, 1997)

A enzima FolM consegue sintetizar tetrahidromonapterina (H4-MPt) através da redução do substrato dihidromonapterina (H2-MPt) (Pribat *et al.*, 2010), esta é uma enzima multifuncional que também catalisa a redução de dihydrofolato, assim como o análogo funcional FolA. Os prováveis resíduos envolvidos na transformação química de dihidromonapterina por FolM são possivelmente os resíduos conservados na família SDR (*Figuras S1*), estes resíduos estão envolvidos na atividade enzimática das proteínas DHPR e são amplamente conservados nos representantes desta Família (Jomvall *et al.*, 1995, Gilard *et al.*, 2003).

Possivelmente, a redução do substrato dihidromonapterina ou dihydrofolato mediada por FolM envolve os mesmos resíduos, devido à redundância estrutural

entre estes substratos. A atividade contra o substrato dihidromonapterina é 16 vezes mais alta (Pribat *et al.*, 2010), o que indica que FoIM têm maior especificidade para dihidromonapterina.

Lockyer e colaboradores (1987) (Lockyer *et al.*, 1987) sugeriram que as atividades enzimáticas análogas de DHPR e DHFR (FoIM) contra pteridinas podem representar homologia vestigial, pois teriam divergido a um ponto onde a similaridade de sequência não pode ser demonstrada estatisticamente. Alternativamente, essa analogia funcional pode refletir um evento de evolução convergente.

Os dados sugerem que diferentes formas análogas podem estar sujeitas a diferentes pressões seletivas (por exemplo, antibióticos), onde uma forma estrutural pode ser mais importante que sua forma estrutural alternativa, e fornecer uma vantagem adaptativa para o organismo possuidor destas enzimas análogas.

As predições *in silico* indicam que a proteína Fola é instável em condição ambiente (Tabela S7), o qual possivelmente é mais um dos motivos pelo qual a forma alternativa FoIM é mantida no genoma, mesmo possuindo baixa atividade para o substrato dihidrofolato, pois a síntese de tetrahydrofolato é essencial para o crescimento de *E. coli*.

Como dito anteriormente, Fola tem alta afinidade (baixo valor de  $K_m$ ) para o dihidrofolato em comparação com FoIM, contribuindo significativamente para a síntese de dihidrofolato na célula. Isto indica que o ajuste no  $K_m$  pode ser uma resposta adaptativa da célula devido à presença seletiva para a síntese de folato. Por outro lado, sua forma análoga FoIM possui afinidade para mais de um substrato, e conseqüentemente a pressão seletiva para esta enzima em direção à síntese do substrato dihidrofolato seria menor.

Uma análise com o programa DISOPRED (Protein Sequence Analysis Workbench) revelou que FoIM possui uma região não estruturada no sítio de ligação com o cofator NADPH (Figuras S1). Sugerimos que isto pode ser um dos motivos da menor afinidade desta enzima com o substrato dihidrofolato, pois tal flexibilidade estrutural no sítio alostérico pode influenciar na ligação da enzima com o substrato, permitindo a ligação de moléculas menores. Esta região desestruturada também é

encontrada na região de ligação ao cofator (N-terminal) nos homólogos pteridinas redutases de *L. donovani* (Keri *et al.*, 2010) e *Trypanosoma brucei* (Dawson *et al.*, 2010) que também catalisam a síntese de dihidromonapterina, tal porção é um crítico componente do sítio ativo (*Figuras S1*).

FoIA pode ser inibida pelos antibióticos trimetropim e metotrexato enquanto FoIM é inibida apenas por metotrexato (Giladi *et al.*, 2003). Na presença de *trimetropim* a concentração de dihidrofolato pode alcançar um nível pelo qual as enzimas com baixa atividade de dihidrofolato redutase e insensibilidade para trimetropim são suficientemente ativas para sustentar o crescimento bacteriano (Giladi *et al.*, 2003). Outro fato a ser relatado é que formas alternativas da dihidrofolato redutase estão envolvidas na resistência contra metotrexato em *L. major*, fornecendo uma via de escape para a síntese de folato (Nare *et al.*, 1997).

Na presença de antibióticos que inibem a atividade enzimática de FoIA, a atividade rudimentar de FoIM pode significar uma rota alternativa para a síntese de tetrahydrofolato. A dihidrofolato redutase é estudada como alvo para drogas em bactérias (Kumar *et al.*, 2012), parasitas (Choowongkomon *et al.*, 2010) e tumores (Askari & Krajinovic, 2010). O aumento da expressão gênica do gene *foIA* através de mutações únicas ou múltiplas no seu promotor pode conferir resistência contra trimetropim e metotrexato (Toprak *et al.*, 2011, Maskell *et al.*, 2001). Com isto, sugerimos que o aumento do número de cópias de *foIA* nas linhagens causadoras de surtos endêmicos (Figura 6) pode prover uma vantagem adaptativa para estas bactérias na resistência contra estes antibióticos.

Em *Streptococcus agalactiae*, as amplificações gênicas dos genes requeridos para a biossíntese de tetrahydrofolato levaram ao surgimento de resistência para trimetropim e sulfonamida (Brochet, *et al.*, 2008).

A cópia adicional de *foIA* está localizada em uma distinta região do genoma onde vários genes vizinhos codificam transposases e uma integrase, o que sugere que este gene possivelmente surgiu de um evento de transposição no genoma destas linhagens (Fig. 7). Como dito anteriormente, a amplificação do gene *foIA* em *E. coli* K12 foi relatada em um estudo experimental populacional envolvendo o crescimento da bactéria sobre diferentes pressões seletivas com o antibiótico trimetropim, onde observou-se um aumento na resistência bacteriana a este

antibiótico (Toprak et al., 2011). Por isto, sugerimos que a presença de uma cópia adicional de uma enzima chave e essencial para a sobrevivência de *E. coli* como Fola pode aumentar a taxa de *fitness* (adaptação e sobrevivência) em condições de tratamento com antibióticos.

Alguns estudos de genômica funcional por meio de análise do fenótipo do mutante para o gene *folA* indica que este é um gene essencial para *E. coli* em crescimento anaeróbico em meio de cultivo mínimo (Gerdes et al., 2003, Baba et al., 2006), Nichols et al., 2011); o mesmo é observado no banco de dados PEC ([www.shigen.nig.ac.jp/ecoli/pec](http://www.shigen.nig.ac.jp/ecoli/pec)). Porém, no trabalho de Howell e colaboradores (1988) (Howell, Foster, & Foster, 1988), *E. coli* mutante para *folA* foi capaz de crescer mas exibiu requerimentos adicionais de metionina, glicina, pantotenato, adenina e guanina. Por outro lado, Ahrweiler e Frieden (1988) (Ahrweiler & Frieden, 1988) reportaram que o mutante para *folA* não necessitou de requerimentos adicionais; provavelmente tais diferenças são devidas aos diferentes aparatos metabólicos entre distintas linhagens de *E. coli*. De fato, a análise das redes de interação de Fola (Fig. 12) e FolM (Fig. 13) aponta para uma maior *betweenness* de Fola em relação a FolM. Por outro lado, esta análise mostra adicionalmente que a maioria das proteínas com as quais FolM interage também o faz com Fola, que ainda interage com várias outras proteínas, o que está de acordo com o relatado em relação à *betweenness* e essencialidade de Fola.

A forma estrutural FolM não é necessária para a sobrevivência de *E. coli*, mas pode atuar como uma forma alternativa para a síntese de tetrahydrofolato no mutante de *E. coli* com a adição de timidina no meio de cultivo (Giladi et al., 2003). Os dados de *fitness* mostram que o mutante *folM* ainda apresenta taxa de *fitness* positiva quando tratado com concentrações crescentes dos antibiótico *trimetoprim* e *sulfamethizole*, ambos potentes inibidores da atividade enzimática mediada por Fola (Figura 9). Possivelmente, a ausência de *folM* induz uma maior expressão de *folA*, resultando em uma menor sensibilidade ao *trimetoprim*. Um resultado similar foi observado por Girgis e colaboradores (2009), o que reforça nossa hipótese de que o aumento do número de cópias de Fola pode prover uma vantagem adaptativa para *E. coli*, como visto nos recentes surtos causados pelas linhagens O104: 2009 EL-2050, O104: 2009 EL2071 (Geórgia) e O104: 2011-3493 (Alemanha).

#### 4.4 Crossover junction endodeoxirribonuclease (EC 3.1.22.4): *rusA*, *ruvC*

Para esta atividade enzimática foram identificados os análogos funcionais *rusA* e *ruvC*. As enzimas RusA e RuvC atuam no processo de recombinação e reparo de DNA através da resolução da Junção de *Holliday* formada durante este processo (Mahdi *et al.*, 1996, Yamada, *et al.*, 2004). Identificamos uma expansão do número de cópias de *rusA* em linhagens patogênicas de *E. coli* enterohemorrágica (EHEC) como, por exemplo, nas linhagens O26: H11 11368, O157: H7 TW14359 possui 8 cópias e O157: H7 EC4115 que possui 9 cópias de *rusA* (Figura 13).

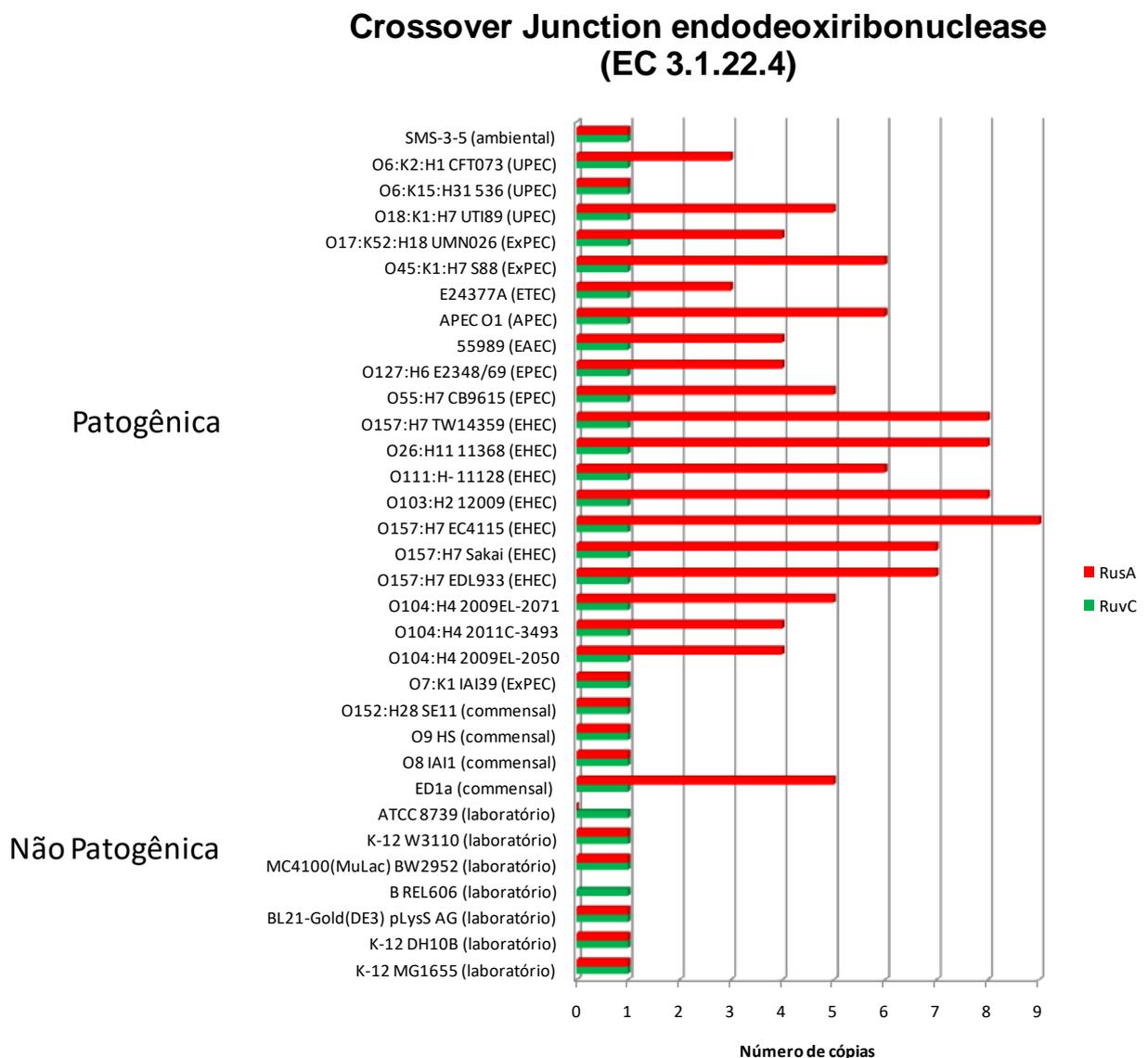
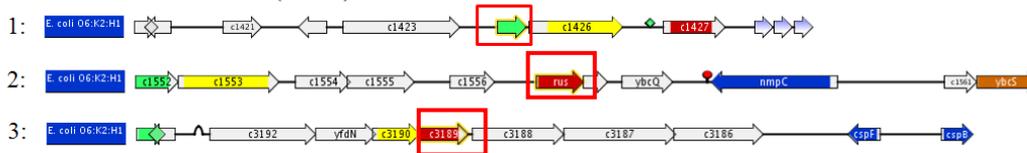


Figura 13: Número de cópias de *rusA* e *ruvC*. EHEC: Enterohemorrágica, EPEC: Enteropatogênica, ExPEC: Estraintestinal, ETEC: Enterotoxigênica.

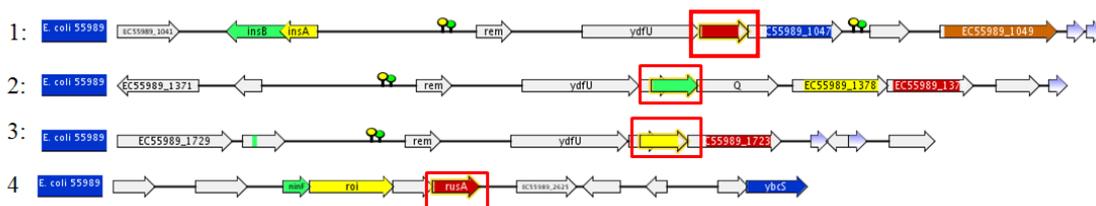
As linhagens comensais e de laboratório possuem apenas 1 cópia de *rusA*, exceto a linhagem Ed1a que possui 5 cópias de *rusA* e ATCC8739 que não possui tal forma estrutural. O número de cópias de *rusA* oscilou entre 1 e 9 cópias por genoma.

Na linhagem O6: K2:H1 CFT073 foram identificadas 3 cópias do gene *rusA*, *upstream* a uma das cópias de *rusA* (lócus: c1556) foi identificado uma sequência que codifica a proteína de inserção *transposase IS629* (lócus c1553) (Figura 14).

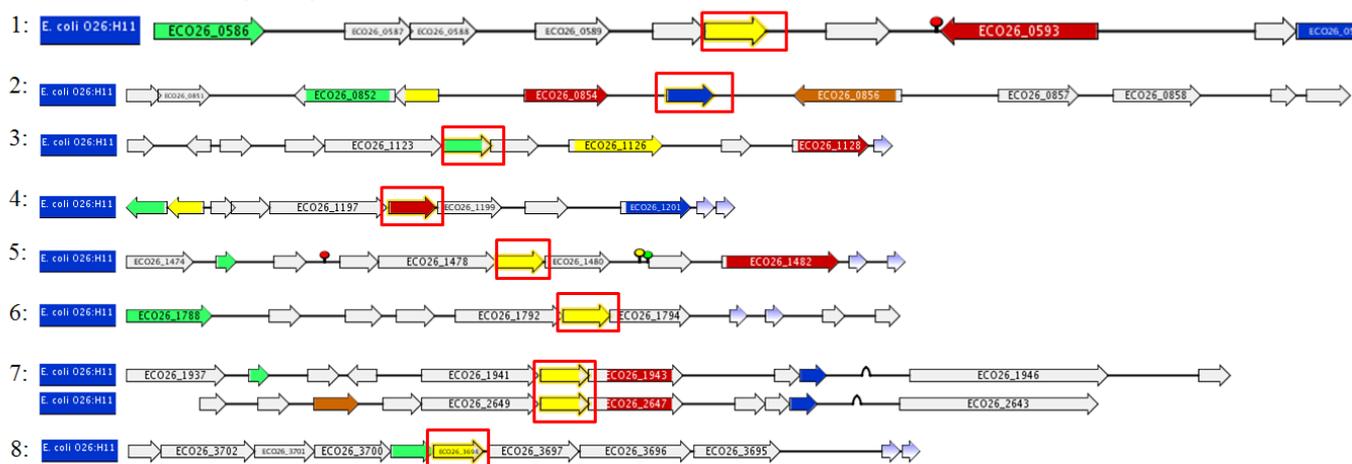
*E. coli* O6:K2:H1 CFT073 (UPEC)



*E. coli* 55989 (EAEC)



*E. coli* O26:H11 11368 (EHEC)



*ruvC* (a), Operon *ruvC*(b), Operon *ruvAB* (c) em *E. coli* K12

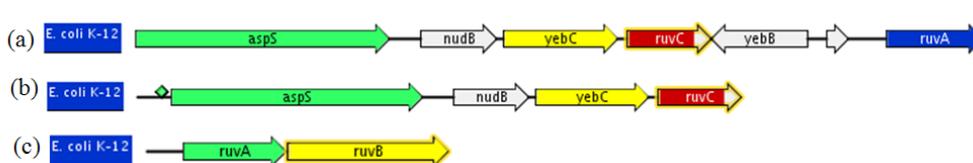


Figura 14: Contexto genômico das diferentes cópias do gene *rusA* em diferentes linhagens de *E. coli*. O contexto genômico de *ruvC* encontra-se na parte de baixo da figura. A legenda completa com a descrição de todos os genes encontra-se na Figuras S2.

Na linhagem 55989 (EAEC) foram identificadas quatro cópias de *rusA* (Figura 14), *upstream* à sequência de uma das cópias de *rusA* (locus: EC55989\_1046) foram identificadas sequências que codificam as proteínas IS1 *repressor protein* InsA (locus: EC55989\_1043) e IS1 transposase InsAB (locus: EC55989\_1042) (anexos). Foi também observado, *downstream* ao gene *rusA*, sequências que codificam tRNAs (loci EC55989\_tRNA15 e EC55989\_tRNA16), *InsA* e *InsAB* são transposons ativos em bactérias; IS1 é responsável por vários rearranjos genômicos, deleções e cointegração de plasmídeos (Machida & Machida, 1989).

Na linhagem O26: H11 11368 (EHEC) foram identificadas 9 cópias de *rusA* e tanto *upstream* ou *downstream* em relação a vários destes genes foram identificados genes que codificam proteínas de profagos, pseudogenes e sequências de inserção (anexos 3).

Nas linhagens O104: H4 str. 2009 EL-2050, O104: H4 2011C-3493, O104: H4 2009EL-2071 e O81 ED1a (comensal) identificamos que uma das cópias de *rusA* está localizada *downstream* ao gene *lexA* e que a linhagem O45: K1: H7 S88 possui três cópias do gene *lexA*, sendo duas localizadas a *downstream* aos genes *rusA* (locus: ECS88\_0550 e ECS88\_2899). Nas demais linhagens estudadas neste trabalho, como, por exemplo, K12, *lexA* está em um locus onde *rusA* não está presente; estas linhagens possuem apenas uma cópia de *lexA*. As cópias do gene *lexA* não localizadas próximas a *rusA*, assim como seu contexto genômico, também foram identificadas nas linhagens O104: H4 str. 2009 EL-2050, O104: H4 2011C-3493, O104: H4 2009EL-2071 e O81 ED1a (*Figuras S2*).

#### **4.4.1 Análise da taxa de sobrevivência de mutantes (*E. coli*) *rusA* ou *ruvC* quando submetidas a diferentes estresses ou compostos químicos**

Para tentar elucidar o papel biológico de *rusA* em *E. coli* obtivemos através do banco de dados PortEco os dados das taxas de adaptação e sobrevivência de mutantes de *E. coli* para os genes *rusA* e *ruvC* durante o tratamento com inibidores de síntese da parede celular, da membrana e com substâncias que causam danos no DNA.

Como esperado, *E. coli* BW25113 mutante para *ruvC* apresenta uma taxa de *fitness* negativa quando tratada com diferentes substâncias que causam danos ao DNA. Em contraste, o mutante *rusA* de *E. coli* BW25113 ainda apresenta um valor de *fitness* ligeiramente positivo em alguns tratamentos com componentes que causam danos ao DNA, exceto com brometo de etídio (TabelaS4).

Observamos que o mutante *rusA*, quando tratado com vancomicina, dibucaína, Chir-090 e bacitracina, apresenta taxa de *fitness* negativa, e em contraste o mutante *ruvC* apresenta uma taxa de *fitness* ligeiramente positiva em algumas condições; sua taxa de *fitness*, mesmo quando negativa, ainda é normalmente maior que a do mutante para *rusA* (TabelaS4).

Através do banco de dados PortEco, observamos que a taxa de *fitness* observada pelo fenótipo de crescimento do mutante *rusA* tem um valor de coeficiente de correlação de 44% com o mutante para o gene *lpxC*. Os mutantes *rusA* e *lpxC* apresentaram uma taxa de *fitness* similar quando ambos foram tratados com os mesmos compostos que atuam na inibição da síntese de membrana, parede celular ou dano ao DNA. O gene *lpxC* codifica a UDP-3-O-acil-N-acetilglicosamina deacetilase (LpxC), esta enzima hidrolisa UDP-3-O-(3-hidroximiristoil)-N-acetilglicosamina em UDP-3-O-(3-hidroximiristoil) glicosamina e acetato; este é o primeiro passo na biossíntese do lipídio A, que é o componente básico do lipolissacarídeo (LPS) da membrana extracelular (Young & Silver, 1991).

Correlações mais fortes (entre um total de 323 compostos químicos e componentes estressantes) foram obtidas para os mutantes *ylcG* (67%), *ninE* (67%) e *asmA* (48%). A proteína NinE também é codificada dentro do profago DLP12, YlcG é uma proteína que é expressa na fase estacionária (Hemm *et al.*, 2008), mas seu papel biológico ainda não foi elucidado; AsmA é uma proteína envolvida na montagem de proteínas, localizada na membrana extracelular. O mutante *ruvC* de *E. coli* é condicionalmente letal na presença de piocianina (Nichols *et al.*, 2011), tal substância é um fator de virulência da bactéria *Pseudomonas aeruginosa* (Muller, 2002)

#### 4.4.2 Comparação estrutural entre RusA e RuvC através de modelagem comparativa e identificação dos resíduos do sítio ativo e sítio alostérico

Os análogos RuvC e RusA (Crossover junction endodeoxiribonuclease) estão envolvidos na resolução da junção de Holliday durante a recombinação homóloga em *E. coli*. A estrutura da proteína RuvC foi classificada no SCOP com o enovelamento “Ribonuclease H-likemotif [53066]” e “Superfamília Ribonuclease H-like [53098]” (Tabela S2)

RuvC é uma proteína dimérica dependente de  $Mg^{2+}$  para a catálise e sua estrutura é constituída de cinco  $\alpha$ -hélices ( $\alpha A$ ,  $\alpha B$ ,  $\alpha C$ ,  $\alpha D$  e  $\alpha E$ ) e uma mistura de folhas- $\beta$  com cinco fitas ( $\beta 1$ ,  $\beta 2$ ,  $\beta 3$ ,  $\beta 4$  e  $\beta 5$ ) (Ariyoshi *et al.*, 1994). A resolução da junção por clivagem de RuvC ocorre em sequências simetricamente relacionadas dentro do grupo de homólogos pareados, especificamente ao lado do resíduo 3'-timina, na maioria dos casos em sequências contendo 5'-TT ↓GC-3', e em menor número 5'-GT ↓CC-3' e 5'-TT ↓GT-3' (R. J. Bennett, Dunderdale, & West, 1993)

O análogo RusA foi classificado no banco de dados do SCOP com o enovelamento “Bacillus chorismate mutase-like” [55297] e Superfamília “Holliday junction resolvase RusA [103084]” (Tabela S2). É uma proteína dimérica com estrutura N- $\beta 1$ - $\alpha 1$ - $\beta 2$ - $\alpha 2$ - $\beta 3$ - $\beta 4$ -C (Rafferty *et al.*, 2003).

A especificidade de clivagem de *RusA* em sequências homólogas pareadas é diferente de RuvC, pois a clivagem ocorre na região 5' de CC-dinucleotídeos. Os resíduos do sítio ativo de RuvC e RusA envolvidos na clivagem da ligação fosfodiéster do fosfato do DNA são mostrados nos anexos (*Figuras S1*). A clivagem da sequência de DNA envolve um ataque nucleofílico de uma molécula de água e os resíduos Lys107 em RuvC e Lys76 em RusA podem estabilizar um estado de transição negativo gerado durante a hidrólise da ligação do fosfato (Rafferty *et al.*, 2003). A convergência dos resíduos do sítio ativo nos análogos RuvC e RusA (*Figuras S1*) e os dados da literatura indicam que os mecanismos de transformação química destes análogos são similares.

#### 4.4.3 Análise dos dados e literatura

Em vários organismos, a replicação seguida por recombinação homóloga é o principal processo biológico responsável por gerar a diversidade genética. Durante este processo o genoma é remodelado através da mistura de genes entre os cromossomos homólogos e também ocorre a reparação de moléculas de DNA danificadas por agentes externos tais como radiação e produtos químicos (West, 1997) A recombinação homóloga envolve a interação e pareamento entre dois segmentos de DNA idênticos ou similares um a outro, cada fita de DNA parental fica pareada à fita homóloga do DNA parental oposto, em seguida o DNA é clivado e ocorre à troca de fitas de DNA entre as sequências homólogas através da formação de um intermediário heteroduplex e finalmente ocorre a separação das fitas recombinadas. Este mecanismo de recombinação genética que envolve a formação de uma estrutura heteroduplex foi proposto por Robin Holliday em 1964 e denominado *crossover*, mas agora a estrutura formada é mais conhecida como junção de Holliday (Watson, & Ho, 2003)

Durante a recombinação em *E. coli* K12 a proteína RecA catalisa a reação de troca de fitas entre as sequências homólogas de DNA e depois o tetrâmero RuvA se liga especificamente na junção de Holliday e guia a ligação da proteína RuvB formando um complexo RuvAB; em seguida, RuvC também se liga formando o complexo RuvABC, a enzima RuvB é uma ATPase que atua como um motor molecular que possibilita a migração do complexo RuvAB na fita de DNA (Bradley *et al.*, 2011), RuvC atua na resolução da junção através de uma clivagem endonucleotídica dependente de metal (Shinagawa & Iwasaki, 1996)

Os genes *ruvA* e *ruvB* constituem um único operon que é regulado negativamente pelo repressor LexA (produto de *lexA*) (Lloyd, 1991); a expressão deste operon é induzida em resposta a danos causados no DNA (Benson *et al.*, 1998). O produto dos genes *ruvA*, *ruvB* e *ruvC* é necessário para a sobrevivência de *E. coli* após o tratamento com luz ultravioleta (Sharples *et al.*, 1990). Os mutantes de *E. coli* para os genes *ruv* exibem um fenótipo de aumento de sensibilidade para radiação ionizante (UV), mitomicina C e em baixas doses de tratamentos com compostos tóxicos que danificam o DNA; estas bactérias exibem viabilidade

reduzida e formam longos filamentos não septados (Otsuji, Iyehara, & Hideshima, 1974).

As linhagens comensais e de laboratório estudadas neste trabalho possuem apenas uma cópia de *rusA* e *ruvC* com exceção da linhagem comensal *E. coli* O81 ED1a, que possui 5 cópias de *rusA* (Figura 13). Esta linhagem é específica de humanos, tendo sido isolada nas fezes de um indivíduo saudável na França em 2000; aparentemente, a frequência desta cepa está aumentando em seres humanos (Clermont *et al.*, 2008). A aquisição de genes de virulência por meio de transferência gênica horizontal durante a evolução de microorganismos pode alterar seu estilo de vida, de uma forma comensal para patogênica. A linhagem O81 ED1a está filogeneticamente mais relacionada com a linhagem Uropatogênica O6: K2: H1 CFT073 (Touchon *et al.*, 2009). Podemos especular que a perda de genes de virulência pode ser um dos fatores associados à não patogenicidade da linhagem O81 ED1a.

A seleção natural é um processo fundamental que afeta todas as populações de organismos. Em casos simples, a seleção positiva aumenta a frequência dos alelos em uma população com mutações que conferem uma habilidade individual para sobreviver e reproduzir em um ambiente. Uma possível evidência de seleção positiva é o aumento da frequência de um alelo em uma população. Já a seleção negativa remove os alelos ou mutações deletérias que diminuem o *fitness* (Lemey *et al.*, 2009). Em outras palavras, em uma determinada população uma mutação benéfica é aquela que aumenta a *fitness* individual em relação ao tipo selvagem, enquanto as mutações deletérias a reduzem (Lemey *et al.*, 2009). Os dados aqui obtidos sugerem que o gene *rusA* pode estar sendo selecionado através de seleção positiva e possivelmente o aumento do número de cópias deste gene em algumas linhagens de *E. coli* está relacionado ao aumento de *fitness* nestas linhagens.

A recombinação genética é uma potente força que dirige a evolução de patógenos microbianos e facilita a aquisição de resistência a antibióticos, determinantes antigênicos e fatores de virulência, que resultam em alteração da patogenicidade (Loughlin *et al.*, 2003). As enzimas *RusA* e *RuvC* são chaves no processo de recombinação e reparo de DNA e atuam na resolução da junção de Holliday formada durante a recombinação genética e reparo de DNA (Mahdi 1996, ,

Yamada *et al.*, 2004). O aumento do número de cópias de *rusA* pode estar relacionado ao aumento da produção de variantes antigênicas e ao aumento da variabilidade genética; conseqüentemente, tais mecanismos podem resultar em bactérias com uma maior virulência ou com inovações em mecanismos de evasão da resposta imune do hospedeiro.

O gene *rusA* em *E. coli* K12 está localizado no fago defectivo DLP12, que é provavelmente derivado de um fago relacionado ao fago 82 e PA-2. O gene *rusA* é pobremente expresso, mas pode ser ativado pela inserção de sequências IS2 ou IS10 *upstream* à sequência promotora da transcrição; tais inserções surgem espontaneamente em linhagens de *E. coli* como supressoras do fenótipo dos mutantes *ruv* (Mahdi *et al.*, 1996). A mutação *rus-1* (localizada na posição entre 580-590 pb, *upstream* ao gene *ompT*) também ativa e suprime o defeito na recombinação em mutantes *ruvAB* ou *ruvABC* (Mandal *et al.*, 1993, Mahdi *et al.*, 1996).

Adicionalmente, uma seleção de linhagens de *E. coli* mutantes para o gene *ruvA* identificou inserções de transposases (Tn1000) *upstream* ao gene *rusA* que suprimem o fenótipo deste mutante (Mahdi *et al.*, 1996). As Tn1000 são transposases conhecidas por carregarem sinais que podem ativar a transcrição de genes adjacentes (Tsai *et al.*, 1987). A expressão de várias cópias do gene *rusA* em plasmídeos suprime o fenótipo de sensibilidade à radiação ultra-violeta (UV) nestes mutantes, tornando tais bactérias resistentes aos efeitos deletérios da radiação, e também resolvendo os problemas na recombinação (Mahdi *et al.*, 1996).

Como dito anteriormente, RusA é uma endonuclease que cliva especificamente as junções de Holliday (Sharples *et al.*, 1994), tal estrutura é formada pela ação da enzima RecA (Robu *et al.*, 2001). A reação catalítica de clivagem mediada por RusA em células irradiadas com UV requer a presença de RecG, uma enzima que catalisa a migração do complexo RuvABC. Algumas linhagens de *E. coli* ainda sobrevivem quando desprovidas de RecG e RuvABC devido à atividade funcional de RusA em um processo que é dependente apenas da enzima RecA (Bolt & Lloyd, 2002).

A transposase IS629 é muito prevalente no genoma de *E. coli* O157: H7 (Rump *et al.*, 2011) e pode possivelmente se replicar através de um mecanismo de

*copy/paste* (Duval-Valentin *et al.*, 2004); esta linhagem possui uma alta taxa de transposição de sequências (Kusumoto *et al.*, 2004). Sabe-se que os rearranjos relacionados às sequências de inserção podem alterar significativamente a virulência e outros fenótipos em *E. coli* (Ooka *et al.*, 2009). Nós sugerimos que a presença de IS629 (lôcus: c1553) *upstream* ao promotor de *rusA* na linhagem O6: K2:H1 CFT073 e em outras linhagens pode possivelmente ativar a expressão de *rusA* nos genomas destas bactérias, pois é um fato conhecido que a integração de elementos móveis podem modificar a expressão gênica em *E. coli* e outros organismos (Schnetzer & Rak, 1992, Mahillon & Chandler, 1998, Saier & Zhang, 2014).

Neste trabalho, identificamos que algumas cópias de *rusA* estão localizadas dentro de profagos, como por exemplo na linhagem O157: H7 EDL93. Nesta linhagem, as diferentes cópias de *rusA* estão localizadas em um dos seguintes profagos: CP-933M, CP-933N, CP-933X, CP-933O (*Figuras S2*). Os dados sugerem que a amplificação de *rusA* nas linhagens patogênicas pode ter ocorrido por eventos de integração de profagos nos genomas e possivelmente também por eventos de transposição. Além disso, verificamos que algumas destas cópias encontram-se inseridas com diferentes sentidos de orientação ou polaridades no genoma (*Figuras S2*). Isto sugere a ocorrência de múltiplos eventos de integração de *rusA* no genoma de algumas linhagens de *E. coli*.

O produto do gene *lexA* controla a expressão de vários genes que controlam a resposta SOS, esta proteína atua como um repressor transcricional através da ligação específica em sequências localizadas nos promotores dos genes regulados por *lexA* (Fernández De Henestrosa *et al.*, 2000). O repressor LexA também regula sua própria síntese (Little, Mount, & Yanisch-Perron, 1981). Quando a molécula de DNA é danificada, a proteína RecA se liga a uma fita de DNA simples para formar um filamento (Cox, 2007, Chen, *et al.*, 2008) este filamento interage com o dímero LexA para ativar sua autoclivagem por um mecanismo alostérico, causando a dissociação de LexA de seu DNA alvo com a ativação do regulon SOS para o reparo do DNA danificado (Kühner *et al.*, 2004, Giese *et al.*, 2008). A flexibilidade conformacional de LexA é o elemento chave para estabelecer uma resposta SOS coordenada (Butala *et al.*, 2011). A proteína LexA possui um domínio N-terminal envolvido na ligação ao DNA via uma região de alfa-hélice-volta-alfa-hélice e um domínio C-terminal envolvido na dimerização e clivagem autocatalítica (Luo *et al.*,

2001). A resposta SOS é essencial para a manutenção do genoma, modula a resposta de resistência a antibióticos e controla a tolerância a multidrogas em subpopulações de células persistentes (Butala *et al.*, 2011).

A presença do gene *lexA upstream* e em um mesmo operon (informação do banco de dados EcoCyc) juntamente com o gene *rusA* dentro de profagos em algumas linhagens de *E. coli* (Figuras S2) sugerem que estes genes são co-transcritos e que possivelmente a expressão do gene *rusA* nestas linhagens pode ser regulada pelo repressor LexA.

A partir desta evidência, sugerimos um circuito adicional e alternativo na resposta de reparo a danos causados no DNA, onde este circuito é possivelmente ativado pela proteína RecA. Neste suposto modelo de ativação, RecA se liga ao DNA danificado clivando LexA de seu operador próximo ao operon, isto ativa a expressão gênica de *lexA*, incluindo o gene *rusA* que é o gene mais próximo *downstream* ao gene *lexA*. Se a expressão gênica e a regulação gênica deste circuito adicional forem confirmadas através de estudos experimentais, isto pode ter implicações nos estudos de resistência microbiana em *E. coli*, pois as linhagens que possuírem *lexA* e *rusA* em um mesmo operon podem apresentar vantagens adaptativas na resposta de reparo, recombinação do DNA e na produção de variantes genéticas de *E. coli*, o que pode levar à modificações nos padrões de resistência a antibióticos e mecanismos de patogenicidade.

Os estudos de *fitness* relativos aos mutantes *rusA* e *ruvC* envolveram mais de 300 compostos (<http://ecoliwiki.net/tools/chemgen/>). Entre esses compostos, os que tiveram maior efeito na taxa de crescimento foram a dibucaína, o composto Chir-090 e a vancomicina. A dibucaína tem propriedades detergentes e atua na membrana plasmática, aumentando a fluidez na membrana e também afetando a força próton motriz mediada pelo gradiente de prótons no envelope celular, o que resulta na diminuição da concentração celular de ATP (Andersen, Holland, & Jacq, 2006). Esta substância é utilizada em laboratório em um método desenvolvido para retirar o flagelo de bactérias (Withan *et al.*, 1978). O composto Chir-090 inibe a biossíntese de lipopolissacarídeos (LPS), o maior componente da membrana externa de bactérias (Bodewits, Raetz, Govan, & Campopiano, 2010), esta substância é o mais potente inibidor da enzima N-acetilglicosaminadeacetilase, uma enzima essencial

para a biossíntese de lipídios em bactérias Gram-negativas (Barb *et al.*, 2007). A vancomicina é um antibiótico que atua na inibição da síntese da parede celular em bactérias Gram-positivas. *Escherichia coli* e outras espécies de bactérias Gram-negativas possuem uma resistência inata a este antibiótico, porque a membrana externa destas bactérias funciona como uma barreira intrínseca à vancomicina (Chen *et al.*, 2011).

Os altos valores de correlação entre *rusA* e *ylcG*, *asmA* e *lpxC* (todas envolvidas em processos relativos à membrana plasmática), juntamente com os resultados individuais da dibucaína, do composto Chir-090 e da vancomicina (todos compostos com efeitos severos na *fitness* e também envolvidos em processos relativos à membrana plasmática) nos leva a especular sobre uma possível participação de *rusA* em algum processo relativo à membrana plasmática, mas isto deve ser melhor investigado.

Na linhagem K12 o gene *rusA* (lócus: Y75\_p0535) está localizado dentro do profago DLP12 (defectivo). Além de *rusA*, os genes *essD*, *rzpD* e *ybcS* (codifica uma lisozima anotada em *E. coli* K12 como o gene *rrrD*) também estão localizados no profago DLP12 e estão envolvidos na formação de biofilmes, a deleção destes genes previne a formação dos mesmos e afetam a integridade da parede celular (Toba *et al.*, 2011a). Por isto, especulamos que *rusA* possa também estar envolvido na formação de biofilmes em *E. coli* e na manutenção da integridade da parede celular. Neste contexto, cabe ressaltar que *ybcS* (*rrrD*) também foi identificada como uma NISE neste trabalho, e também encontram-se amplificadas no genoma de linhagens de *E. coli* patogênicas (Figura 24 e Figuras S1)

Outro dado ainda não mencionado é que alguns estudos de mutante indicam que *rusA* e *ruvC* são essenciais para *E. coli* em crescimento anaeróbico em meio de cultivo mínimo (Gerdes *et al.*, 2003, Baba *et al.*, 2006, Nichols *et al.*, 2011), o mesmo é observado no banco de dados PEC ([www.shigen.nig.ac.jp/ecoli/pec](http://www.shigen.nig.ac.jp/ecoli/pec)).

Mahdi e colaboradores (1996), demonstraram que RuvC e RusA não possuem similaridade entre suas sequências de aminoácidos. Posteriormente, Aravind e colaboradores (2000) sugeriram que a enzima RusA evoluiu independentemente de quaisquer outras resolvases. Em nosso trabalho aprofundamos um pouco mais estes dados, mostrando os distintos enovelamentos

adotados por RusA e RuvC além de identificarmos que estas sequências possuem convergência de resíduos no sítio ativo, o que indica que estas enzimas utilizam um mesmo mecanismo de transformação química (convergência mecanística) (*Figuras S1*).

#### **4.5 Glicerol-3-fosfato 1-O-aciltransferase (EC 2.3.1.15): *plsX*, *plsB* e *plsY***

Uma enzima pertencente à classe transferases, a glicerol-3-fosfato 1-O-aciltransferase está envolvida na biossíntese de lipídeos da membrana plasmática e possui três formas análogas em todas as linhagens de *E. coli* estudadas neste trabalho, ou seja, possui três distintas formas estruturais que catalisam a síntese de 1-acil-sn-glicerol 3-fosfato (*Tabela S2*), todas as linhagens possuem apenas uma forma estrutural enzimática de cada enovelamento. A via metabólica dos glicerofosfolipídeos é mostrada na Figura 15.



metabolismo de glicerol e com a proteína LexA (regulador global da resposta SOS), entre diversas outras (Figura 16).

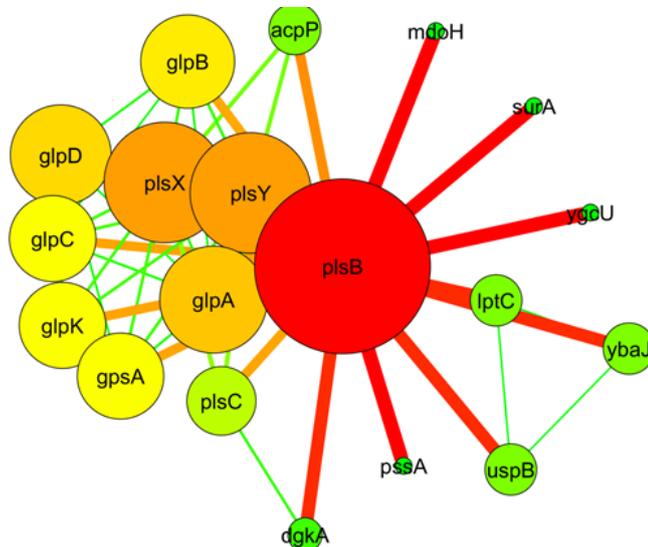
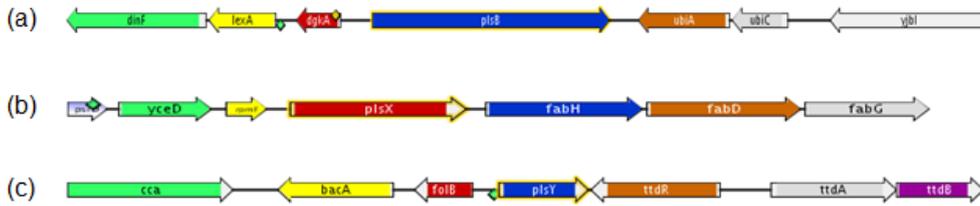


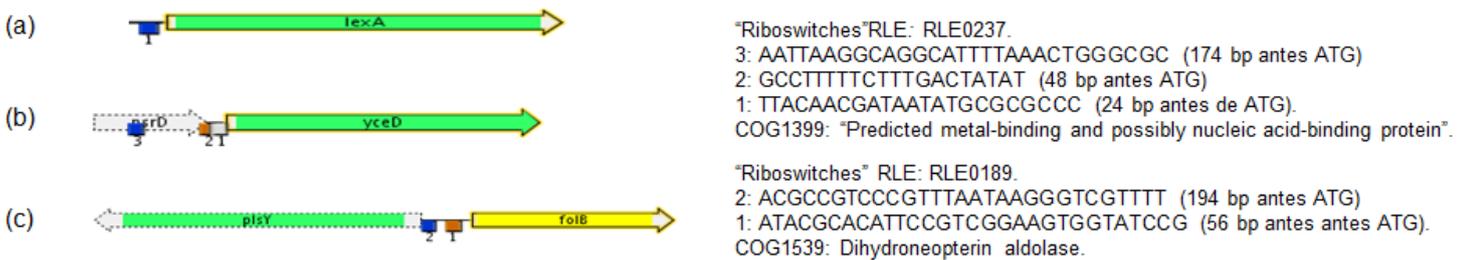
Figura 16. Rede de interação de PlsB. Os dados foram obtidos no banco de dados *STRING* (*STRING*-db.org). Resultado da representação da intensidade dos valores obtidos das medidas grau de interação e *betweenness* através de análises no programa *Network analysis*, implementado no *programa programa Cytoscape* (versão 2.8). O tamanho dos nós e a intensidade das cores na rede representam os valores obtidos de grau de interação e *betweenness*, as cores mais intensas denotam maiores valores. Os análogos *plsX* e *plsY* foram excluídos. A legenda completa com a descrição de todos os genes encontra-se nos (FigurasS3).

O contexto genômico dos análogos funcionais *plsB*, *plsX*, *plsY* obtidos com o *programa GeneContextToolIII* é mostrado na Figura 17. Através da predição do contexto genômico, identificamos que as sequências dos genes *lexA*, *prsB* e *folB*, localizados *upstream* aos genes *plsB*, *plsY* e *plsX*, respectivamente, contêm *riboswitches* em seus RNA mensageiros (Figura 17). Através da análise da rede de interação *plsX* (Figura 18), identificamos que a mesma interage com enzimas envolvidas no metabolismo de lipídios e com proteínas envolvidas na síntese de proteínas ou no metabolismo de ácidos nucleicos.

A: Contexto genômico de *plsB*, *plsX* e *plsY*



B: Presença de *riboswitches* em genes vizinhos aos análogos funcionais *plsB*, *plsX* e *plsY*



Presença de *riboswitches* em genes vizinhos a *plsB*, *plsX* e *plsY*. (a) *lexA*, (b) *yceD*, (c) *foIB*.

Figura 17. A: Contexto genômico de *plsB*, *plsX* e *plsY* em *E. coli* K12. (a) *plsB*, (b) *plsX* e (c) *plsY* B: Presença de *riboswitches* em genes vizinhos a *plsB*, *plsX* e *plsY*. (a) *lexA*, (b) *yceD*, (c) *foIB*.

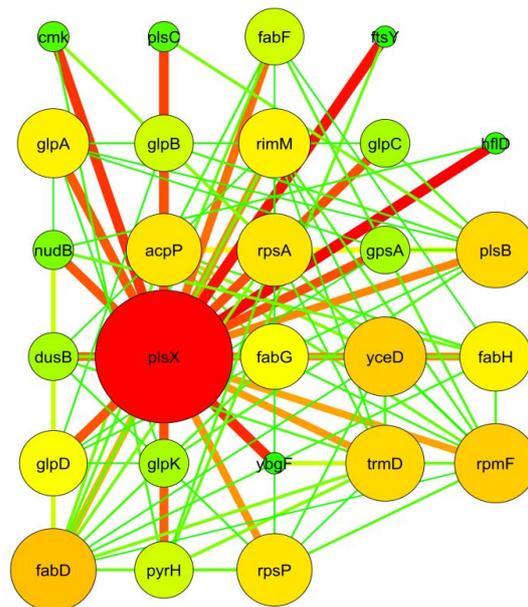


Figura 18. Rede de interação de PlsX. Dados obtidos no banco de dados *STRING* (*STRING*-db.org/). Resultado da representação da intensidade dos valores obtidos das medidas grau de interação e *betweeness* através de análises no programa Network analysis, implementado no programa *Cytoscape* (versão 2.8). O tamanho dos nós e a intensidade das cores na rede representam os valores obtidos de grau de interação e *betweeness*, as cores mais intensas denotam maiores valores. A legenda completa com a descrição de todos os genes encontra-se nos anexos (FigurasS3).

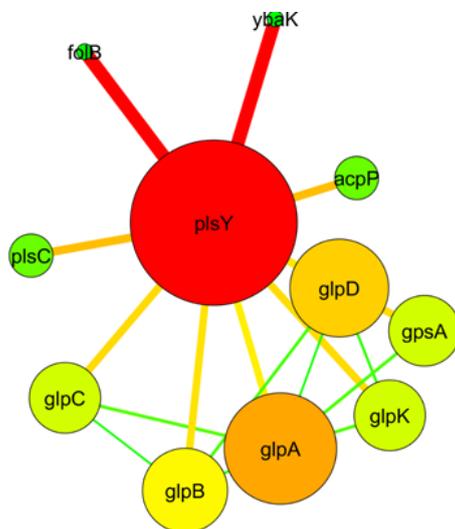


Figura 19. Rede de interação de PlsY. Dados obtidos no banco de dados *STRING* (*STRING*-db.org/). Resultado da representação da intensidade dos valores obtidos das medidas grau de interação e *betweeness* através de análises no programa Network analysis, implementado no programa *Cytoscape* (versão 2.8). O tamanho dos nós e a intensidade das cores na rede representam os valores obtidos de grau de interação e *betweeness*, as cores mais intensas denotam maiores valores. A legenda completa com a descrição de todos os genes encontra-se nos anexos (FigurasS3).

## 4.5.2 Análise da taxa de sobrevivência de mutantes *plsB*, *plsX* e *plsY* quando submetidas a diferentes estresses ou compostos químicos

Para tentar elucidar o papel biológico de *plsX*, obtivemos os dados de *fitness* do mutante para este gene em *E. coli* BW25113 através do banco de dados PortEco e analisamos a taxa de *fitness* durante o tratamento deste mutante com os antibióticos spectinomomicina, eritromicina, *trimetropim* e *sulfamethizole* (Figura 20).

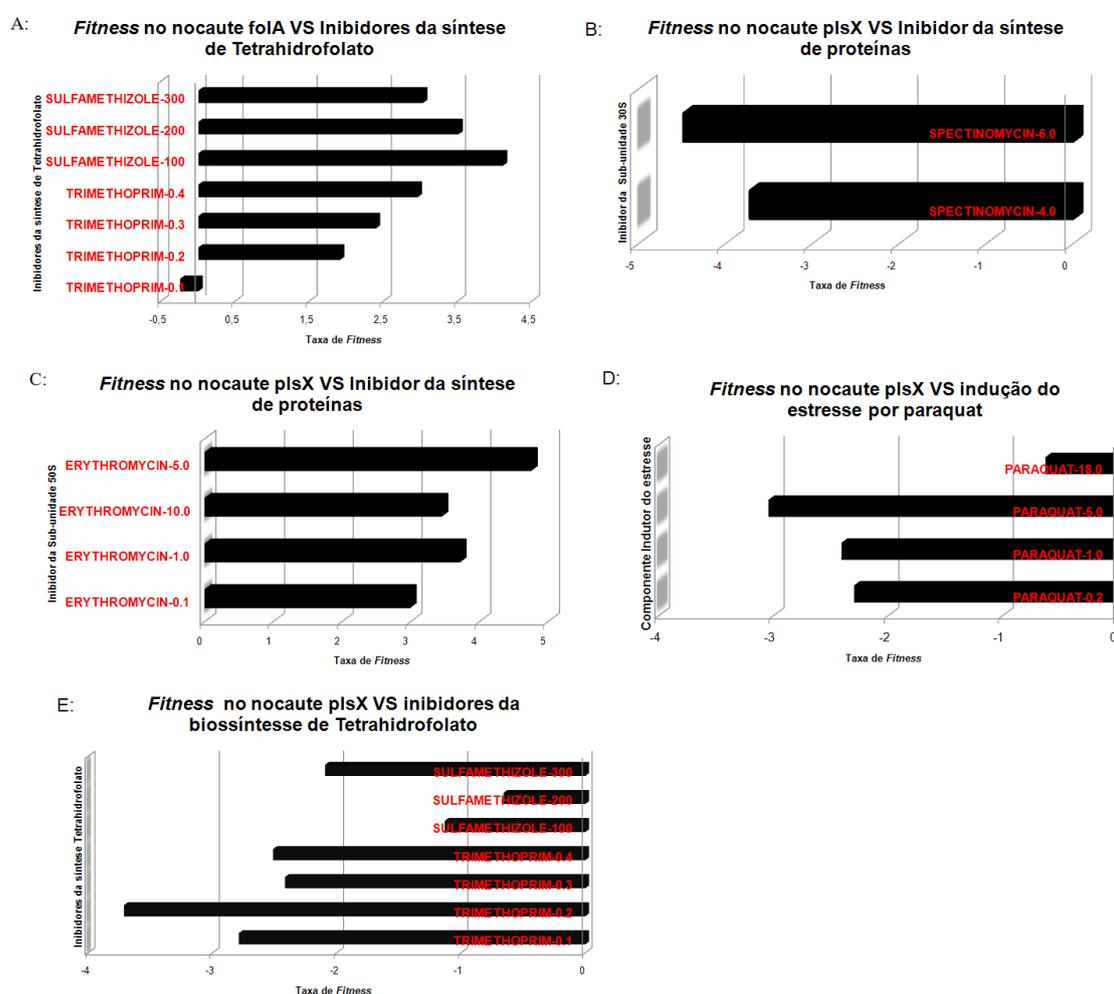


Figura 20: A: Taxas de *fitness* dos mutantes *plsX*, *plsY* e *folA* para *Escherichia coli* BW25113 tratado com inibidores da síntese de proteínas e indutor de estresse Paraquat. A: Taxas de *fitness* no mutante *folA* com inibidores da síntese de tetrahidrofolato. B: Taxas de *fitness* no mutante *plsX* tratado com spectinomomicina. C: Taxas de *fitness* no mutante *plsX* com eritromicina. D: Taxas de *fitness* no mutante *plsX* tratado com o agente indutor de estresse Paraquat. E: Taxas de *fitness* no mutante *plsX* tratado com o inibidor da síntese de tetrahidrofolato. Dados obtidos do banco de dados PortEco.

### 4.5.3 Análise dos dados e literatura

O primeiro passo da via metabólica dos glicerofosfolípídeos é a síntese de sn-glicerol 3-fosfato (G3P) proveniente de dihidroxiacetona fosfato, catalisada pela enzima GpsA (Glicerol-3-fosfato dehidrogenase (NAD (P) (+))). No próximo passo da via, a enzima PLSB (acil-ACP: sn-glicerol-3-fosfato 1-O-aciltransferase) liga o primeiro ácido graxo com G3P (glicerol 3-fosfato) para formar monoacil-G3P (ácido lisofosfatídico). Esta mesma reação pode ser catalisada pelos análogos funcionais PlsX, PlsB e PlsY. A enzima PlsB é pobremente conservada fora do clado das Alfa e Gama proteobactérias e micobactérias (Cronam *et al.*, 2003). O próximo passo é realizado pela enzima PlsC (acil-ACP: 1-acil-sn-glicerol-3-fosfato 2-O-aciltransferase) que catalisa a ligação do segundo ácido-graxo para formar um intermediário de ácido fosfatídico que vai ser convertido em CDP-di-glicerídeo (CDP-diacilglicerol) pela enzima CDP-di-glicerídeosintase.

A molécula de CDP-di-glicerídeo pode seguir duas distintas rotas metabólicas: pode ser utilizada para a síntese de fosfatidiletanolamina, passo catalisado pela enzima PssA, ou pode ser utilizada para a síntese de fosfatidil glicerol e cardiolipina pela enzima PgsA (Cronan, 2003). As enzimas PssA e PgsA também são análogos funcionais (Omelchenko *et al.*, 2010), mas não identificamos este caso de analogia em nosso trabalho. Os análogos PssA e PgsA fornecem rotas alternativas para a síntese de componentes da membrana plasmática, sendo um claro exemplo de que os análogos intragenômicos podem ser utilizados como rotas alternativa em uma via metabólica.

Os estudos de mutantes (Gerdes *et al.* 2003, Baba *et al.*, 2006, Nichols *et al.*, 2011) e dados obtidos do banco de dados PEC indicam que *p/sB* é essencial para a sobrevivência de *E. coli*. A enzima PlsB é localizada na membrana interna e integrada na membrana, PlsY é localizada integrada na membrana plasmática e PlsX é localizada no citoplasma (Tabela S8).

De acordo com a literatura, a enzima PlsB interage com a multi-subunidade glpABC (Ehrmann *et al.*, 1987) e com ACP, YbgC e PssA formando um complexo na membrana interna (Gully e Bouveret *et al.*, 2006). A proteína GlpC é responsável por

ancorar as proteínas GlpA e GlpB na membrana (Cole *et al.*, 1988). Estudos em *E. coli* indicam que PgsA, análogo funcional de PssA é essencial para a sobrevivência da mesma (TabelaS5), mas o análogo funcional PgsA não é essencial. As enzimas PlsB, GlpABC, GlpD e LexA estão envolvidas na formação de células persistentes que têm como característica a resistência a multidrogas (Spoering *et al.*, 2006).

A superexpressão de glpD resulta em um aumento na formação de células persistentes, enquanto sua deleção não diminui a persistência em fase exponencial, mas elimina a maioria da persistência em fase estacionária. Isto indica um envolvimento do metabolismo de glicerol na formação de células persistentes (Spoering *et al.*, 2006). Populações de bactérias são capazes de produzir células dormentes que são resistentes para todos os antibióticos atualmente em uso e a população de células persistentes, especialmente em biofilmes, protege as células do sistema imune; em *E. coli* as células persistentes são formadas e aumentam acentuadamente na fase estacionária atingindo 1% de frequência na população (Spoering *et al.*, 2006).

Com relação à função das proteínas que interagem com PlsB, YciS é uma proteína localizada na membrana interna com dois domínios transmembrana e com uma região C-terminal localizada no citoplasma (Daley *et al.*, 2005); dados de Klein e colaboradores (2014) sugerem que esta proteína coordena a montagem de proteínas envolvida na síntese de LPS na membrana plasmática e regula a reciclagem de LpxC (N-acetilglicosaminadeacetilase) para assegurar o balanço da biossíntese de LPS e fosfolipídios; SurA é uma peptidil-prolil cis-transisomerase (também identificada em nosso trabalho como uma NISE, TabelaS2) necessária para o correto enovelamento de proteínas acopladas na membrana extracelular, sendo essencial para a formação de biofilmes (Niba *et al.*, 2007) e a formação de células de persistência (Hansen *et al.*, 2008); LptC é um componente do sistema de transporte de lipopolissacarídeos, essencial para o crescimento de *E. coli* (Baba *et al.*, 2006, Sperandeo *et al.*, 2006); Tus é uma proteína que bloqueia a função da helicase, impedindo o progresso da forquilha de replicação durante a duplicação do DNA (Henderson *et al.*, 2001) reprimindo a replicação (Hill *et al.*, 1989); UspB (*universal stress protein B*) é uma proteína de resposta a diferentes condições de sobrevivência e estresses (Farewell *et al.*, 1998); YbaJ é uma antitoxina altamente induzida durante a formação de biofilmes, que possivelmente forma um complexo

com a toxina Hha para inibir a ação desta proteína, a qual é responsável pela diminuição da formação de biofilmes e lise celular (García-Contreras *et al.*, 2008); MdoH (também denominada OpgH) é uma glicosiltransferase que catalisa a síntese de oligossacarídeos de membrana localizados no espaço periplasmático (Weissborn *et al.*, 1984).

Modelos matemáticos sugerem que a interação conjunta de múltiplos sistemas toxina e antitoxina (TA) podem regular o fenótipo de células de persistência (Fasani & Savageau, 2013). Interessantemente, todos os componentes da rede PlsB que mapeamos estão envolvidos na síntese e transporte de LPS, que são moléculas envolvidas na formação de biofilmes (Nakao *et al.*, 2012) e na formação de células de persistência, reforçando o papel biológico de PlsB e seus interatores na formação de células de persistência e biofilmes.

Nossos dados estão de acordo com a hipótese sugerida por Spoering e colaboradores (2006) sobre o envolvimento do metabolismo de glicerol na formação de células de persistência. Em condições deletérias para *E. coli*, como por exemplo, na presença de antibióticos, a via de resposta de formação de células de persistência é ativada promovendo a adaptação e resistência da bactéria. Modelos matemáticos também mostram que a persistência pode assegurar a sobrevivência e adaptação (*fitness*) em populações de bactérias que vivem em ambientes que mudam frequentemente (Kussell *et al.*, 2005).

A via de ativação de resposta SOS é uma via de sinalização que regula o reparo de DNA e está envolvida na formação de células persistentes; mutantes de *E. coli* com deficiência em genes que participam da resposta SOS (como *lexA*, *recA* ou *recB*) exibem níveis de persistência reduzida quando tratados com o antibiótico ciprofloxacina, sugerindo um mecanismo ativo de formação de células persistentes mediado pela resposta SOS (Dorr *et al.*, 2009).

Os *riboswitches* são sequências regulatórias em RNAs mensageiros que são capazes de se ligar a metabólitos ou íons de metais e regular a expressão do mRNA através da formação de estruturas alternativas em resposta a estes ligantes (Nudler e Mironov 2004; Tucker e Breaker 2005; Winkler 2005). Os *riboswitches* podem detectar a presença de uma variedade de ligantes tais como: íons de magnésio, precursores de ácidos nucleicos, cofatores enzimáticos e resíduos de aminoácidos

(Hammann e Westhof 2007). A presença de *riboswitches* próximos a *plsY* e ao gene *foB* também sugere uma possível ligação entre a enzima glicerol-3-fosfato 1-O-aciltransferase e o metabolismo de síntese de folato.

Nas espécies *Streptococcus pneumoniae* e *Bacillus subtilis* as enzimas PlsY e PlsX catalisam um passo para a formação de fosfolipídeos (Yoshimura *et al.*, 2007). Em *E. coli*, a deleção concomitante de *plsY* e *plsX* é letal (Larson *et al.*, 1984). Yoshimura e colaboradores (2007)(Yoshimura *et al.*, 2007) propuseram que as enzimas PlsY e PlsX atuam no citoplasma regulando a concentração intracelular de acyl-AcP. O gene *plsX* está localizado no genoma de *E. coli* *upstream* aos genes *fabH*, *fabD* e *fabG*, tais genes codificam três  $\beta$ -cetoacil-ACP sintases, enzimas envolvidas na biossíntese de ácidos graxos (Verwoert, *et al.*, 1992).

Com relação ao papel biológico de PlsX, os dados sugerem que esta enzima participa de uma rede biológica que atua na tradução e síntese de proteínas. Com relação à função das proteínas que interagem com PlsX, a enzima DusB é responsável por sintetizar 5,6-di-hidouridina que é uma base nitrogenada modificada encontrada abundantemente em tRNAs (Bishop *et al.*, 2002); TrmD é uma enzima essencial para o crescimento de *E. coli* (Baba *et al.*, 2006), ela catalisa a adição de um grupo metil para G37 de certos tRNAs (Hjalmarsson *et al.*, 1983); RpsP é um dos componentes da subunidade 30S do ribossomo, é essencial para a viabilidade (Person *et al.*, 1995) e pode atuar como uma endonuclease dependente de  $Mg^{2+}$  (Oberto *et al.*, 1996), podendo também se ligar a  $Zn^{2+}$  (Katayama *et al.*, 2002); esta atividade de endonuclease em ribossomos foi descrita há 60 anos por Lehman e colaboradores (1962); RimM é um fator de montagem para a subunidade 30S do ribossomo estabilizando a estrutura do RNA ribossômico (Guo *et al.*, 2013), mutantes para esta proteína exibem uma lenta taxa de tradução (Lovgren *et al.*, 2004); CmK (citidilatocinase) atua na refosforilação de CMP e dCMP (Bertrand *et al.*, 2002); YceD ainda não tem função conhecida; RpsA (também denominada proteína S1) é essencial para *E. coli*, é um componente da subunidade 30S do ribossomo: durante a tradução, S1 se liga à sequência líder do mRNA *upstream* à sequência Shine-Delgarno (Komarova *et al.*, 2002), é também necessária para ligar o tRNA ao ribossomo durante a tradução (Bordeau e Felden, 2002); RpmF (também denominada L32) é um componente da sub-unidade 50S, localizada próximo ao centro da peptidiltransferase (Muralikrishna e Cooperman, *et al.*, 1995); PyrH é uma

enzima essencial envolvida na síntese *de novo* de pirimidinas (Serina *et al.*, 1995); NudB catalisa a remoção hidrolítica de pirofosfato de dihidroneopterina trifosfato para produzir dihidroneopterina fosfato, é uma enzima chave na via de biossíntese de folato: mutantes para este gene exibem uma redução marcante na biossíntese de folato (Gabelli *et al.*, 2007); HflD atua na lise do bacteriófago  $\lambda$  (Kihara *et al.*, 2001); FstY é o receptor para a partícula SRP, a qual atua na inserção de proteínas recém traduzidas na membrana plasmática interagindo com o transportador de translocação SEC (Luring *et al.*, 1994); YbgF é um dos componentes do sistema Tol-Pal que atua na manutenção da integridade da membrana plasmática permitindo a formação de um complexo trans-envelope entre a membrana interna e externa (Walburguer *et al.*, 2002).

Os resultados de interação da rede biológica de PlsX (Figura 18) sugerem uma ligação entre a via de síntese de proteínas e a enzima PlsX que participa do metabolismo de glicerol. Tal associação foi proposta inicialmente por Podkowrov e Larson (1995) e nossos dados de predição computacional e análise de mutantes de *E. coli* dão suporte para uma rede de regulação coordenada entre PlsX e a síntese de proteínas.

A rede de interação do análogo PlsY envolve algumas das proteínas supracitadas e FolB (Figura 19) que é uma importante enzima da via de síntese de folato usada como alvo antibacteriano. Esta enzima catalisa a clivagem de 7,8-dihidro-d-neopterina (DHNP) para formar 6-hidroximetil-7,8-dihidropterina (HP) e glicolaldeído (Wang *et al.*, 2007), catalisando a epimerização reversível da posição 3' de di-hidroneopterina produzindo dihidromonapterina (Haußmann *et al.*, 1998), um substrato para a enzima FolM (dihidrofolato redutase).

O metabólito glicolaldeído pode induzir a ativação de membros do sistema regulon SoxRS (família de ativadores da transcrição) que induzem a expressão das enzimas fumarato hidratase C e nitroreductase A; esta indução é dependente da auto-oxidação de gliceraldeído através de superóxidos ( $O_2^-$ ). Este sistema (SoxRS) induz uma resposta ao estresse que protege a célula contra o estresse oxidativo causado por superóxidos ( $O_2^-$ ) e gás de óxido nítrico dissolvido (Benove Fridovich *et al.*, 2002). A enzima nitroreductase A é induzida em *E. coli* na presença de Paraquat de forma dependente de soxR (Liochev *et al.*, 1999).

Os genes *rpmF*, *plsX*, *fabH*, *fabD* e *fabG* são co-transcritos e organizados em um operon e regulados por um promotor upstream ao gene *rpmF*, sugerindo que isto pode representar um papel na regulação coordenada da síntese de ribossomos e membrana celulares (Podkoryov e Larson *et al.*, 1995). *Upstream* ao gene *fabH* existe um segundo promotor que regula a expressão dos genes *fabH*, *fabD* e *fabG*, localizado dentro da sequência *plsX*. Isto indica que estes dois promotores podem responder a diferentes estímulos (Podkoryov e Larson *et al.*, 1996). De acordo com Zhang e Cronan (1998), 60% da transcrição do gene *plsX* inicia-se no promotor localizado *upstream* ao gene *rpmF*, mas *plsX* têm seu próprio promotor que é suficiente para o crescimento normal de *E. coli*.

A spectinomicina atua na inibição da síntese de proteínas através de sua ligação na hélice 34 do RNA na cabeça do domínio da subunidade 30S do ribossomo, o que resulta em uma mudança na conformação desta subunidade; isto inibe a translocação do tRNA e do mRNA sobre o ribossomo (Borovinskaya *et al.*, 2007). A eritromicina se liga na sub-unidade 50S e inibe a liberação do tRNA no sítio doador do ribossomo, conseqüentemente inibindo a translocação (Tanaka *et al.*, 1973).

O metabólito tetraidrofolato é sintetizado pelos análogos funcionais F<sub>o</sub>IA e F<sub>o</sub>IM e é um precursor da biossíntese de aminoácidos, purinas e pirimidinas que serão utilizados para a síntese de proteínas. Células de *E. coli* tratadas com inibidores da síntese de tetraidrofolato mediada por F<sub>o</sub>IA utilizam como rota de catálise alternativa a enzima F<sub>o</sub>IM. Por isto, a linhagem deficiente para *folA* ainda apresenta uma taxa de *fitness* positiva (Figura 20A). Mas, interessante, observamos que o mutante *plsX* tratado com inibidores da síntese de folato apresentaram uma taxa de *fitness* negativa (Figura 20). Em nossa rede de interação observamos que *plsX* interage com *nudB* (dihidroneopterin trifosfatopirofosfatase); esta enzima catalisa a síntese de di-hidroneopterin fosfato, passo anterior à etapa mediada pela dihidrofolato redutase. Portanto, os dados sugerem um papel adicional de *plsX* na regulação da via metabólica de biossíntese de tetraidrofolato e síntese de proteínas.

Outro fato que concerne o papel biológico de *plsX* é que o mutante para este gene apresenta uma taxa de *fitness* negativa quando tratado com Paraquat (Figura

20D). Nós sugerimos que *p/sX* também participa da resposta SOS, assim como seu análogo funcional *p/sB*, que é essencial e aparentemente está ligado á produção de biofilmes).

Em resumo, os dados indicam uma possível participação de *p/sX* na síntese de proteínas, de *p/sY* no metabolismo de tetraidrofolato e de *p/sB* na formação de biofilmes, evidenciando distintos papéis biológicos destas proteínas. Adicionalmente, a essencialidade de *p/sB*, juntamente com seu possível papel na formação de biofilmes, torna necessária uma maior investigação sobre este gene.

#### **4.6 5'-nucleotidase (EC 3.1.3.5): SurE, UshA, YfbR, YjjG**

##### **4.6.1 Número de cópias de *surE*, *ushA*, *yfbR* e *yjjG***

Nossas análises mostram que a atividade enzimática 5'-nucleotidase possui 4 distintas formas enzimáticas em *E. coli* K12: SurE, UshA, YfbR e YjjG. Em algumas linhagens patogênicas de *E. coli*, observamos um maior número de cópias de *ushA* (Figura 21).

### 5'-nucleotidase (EC 3.1.3.5)

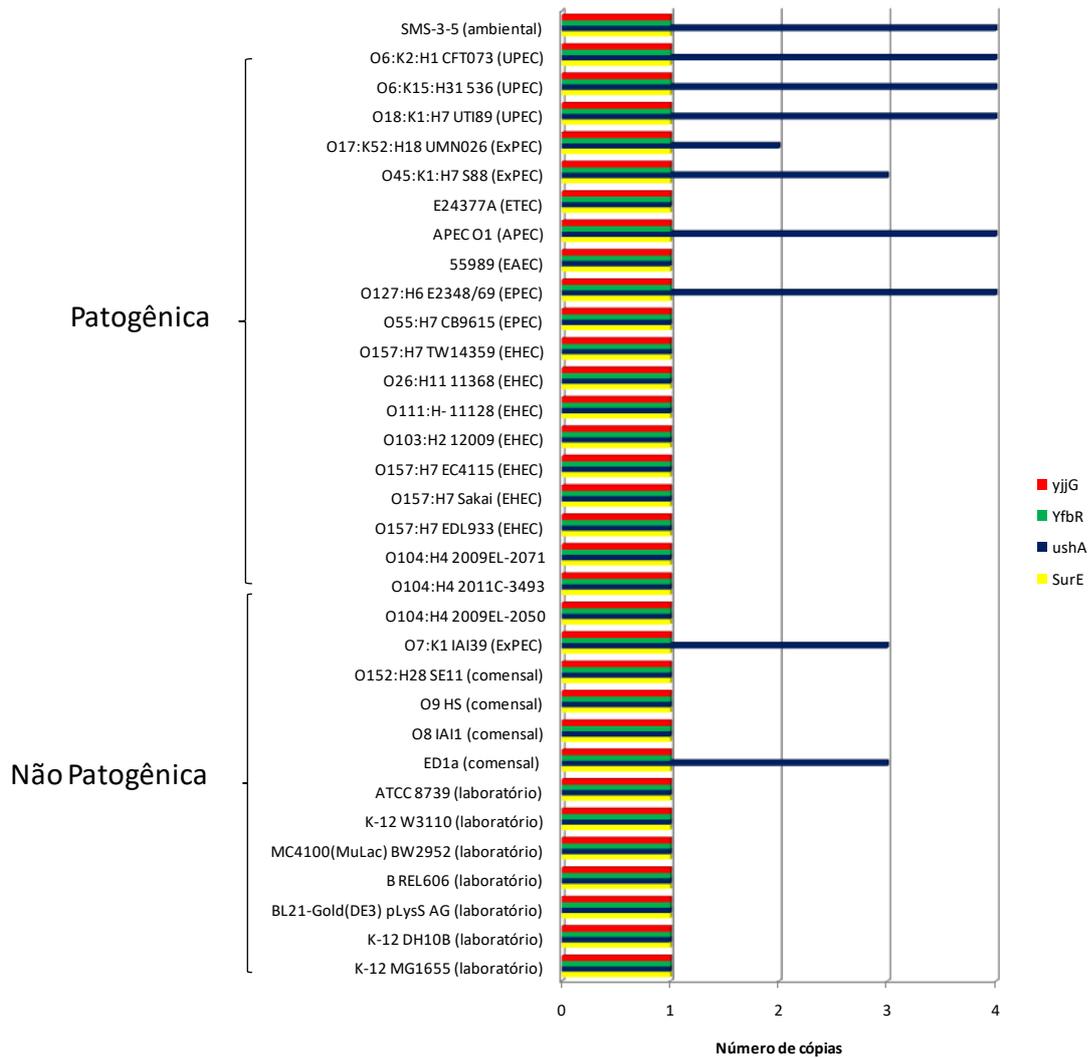


Figura 21: Número de cópias de *surE*, *ushA*, *yfbR* e *yjiG* nas 33 linhagens estudadas. EHEC, Enterohemorrágica, EPEC: Enteropatogênica, ExPEC: Extraintestinal, UPEC: Uropatogênica, APEC: patogênica de aves.

As enzimas 5'-nucleotidasas catalisam a hidrólise ou transferência de um fosfato esterificado para a posição 5' de nucleosídeos mono-fosfato e são as responsáveis pela regulação dos níveis de nucleosídeos e nucleotídeos na célula; elas estão diretamente envolvidas nas vias de biossíntese de purinas e pirimidinas (Bogan, 2010).

Os nucleosídeos e nucleotídeos são os blocos construtores essenciais para a síntese de DNA e RNA, e conseqüentemente essenciais para a existência da vida. A manutenção apropriada da razão de nucleotídeos é crítica para a manutenção e acurácia no reparo e replicação de DNA (Bogan, 2010). Neste contexto, a origem independente de quatro distintas formas de 5'-nucleotidase e a manutenção destas enzimas no genoma de *E. coli* durante sua evolução reflete a importância desta atividade enzimática para a manutenção do metabolismo de nucleotídeos e perpetuação da espécie bacteriana.

Em *Xylella fastidiosa* a 5'-nucleotidase é altamente expressa durante a formação de biofilmes (Santos *et al.*, 2013). Esta atividade foi associada com patogenicidade em *Staphylococcus aureus* (Thammavongsa *et al.*, 2009), *Bacillus anthracis*, *Streptococcus sanguinis* (Fan *et al.*, 2012) e *Haemophilus influenza* (Zagursky *et al.*, 2000)

Em *Staphylococcus aureus*, durante a infecção de células sanguíneas, o aumento da expressão da proteína de superfície AdsA (também uma 5'-nucleotidase) é responsável pelo aumento de adenosina extracelular, um potente imunossupressor (Kim *et al.*, 2011). Observamos (dados não mostrados) que as 5'-nucleotidasas AdsA (S4X8D0) e NucA (Q9X6T9, *Haemophilus influenza*) estão classificadas dentro das mesmas superfamílias no banco de dados SCOP, ou seja, tais proteínas são ortólogas. Nossos dados sugerem que a presença de cópias adicionais de *ushA* em linhagens patogênicas de *E. coli* pode estar relacionada com mecanismos de patogenicidade ou supressão da resposta imune do hospedeiro.

O análogo SurE é codificado por um operon que é regulado durante a fase estacionária pelo fator sigma e têm o papel de aumentar a viabilidade da bactéria durante condições de estresse (Pappachan, Savithri, & Murthy, 2008a). Portanto,

análogos para a atividade 5'-nucleotidase também possuem distintos papéis biológicos.

#### 4.6.2 Comparação estrutural e modelagem comparativa

Como dito anteriormente, a 5'-nucleotidase (EC 3.1.3.5) possui 4 distintas formas enzimáticas em *E. coli*, codificadas pelos genes *surE*, *ushA*, *yfbR* e *yjjG*. Isto proporciona uma oportunidade ímpar de examinarmos a evolução, em termos estruturais, dos sítios ativos desta enzima, uma vez que sabemos que suas estruturas terciárias são distintas.

A enzima YjjG é um membro da superfamília haloácido dehalogenase (HAD)-like hidrolase; a maioria destas proteínas não têm função conhecida. Entre as funções conhecidas destacam-se as atividades de haloácido dehalogenase, fosfonohidrolase, fosfatase, fosfoglucomutase ou ATPase (Proudfoot et al., 2004). Algumas destas proteínas catalisam a defosforilação de pequenas moléculas (fosfoserinas ou fosfoglicolato) ou fatores de transcrição (Proudfoot et al., 2004). A enzima YjjG de *E. coli* é uma proteína dependente de  $Mn^{+2}$  que atua na defosforilação de dUMP, dTMP e UMP. Os resultados de predição de estrutura secundária indicam que Yjj possui 9-12 regiões de  $\alpha$ -hélices e seis regiões de folhas- $\beta$  (dados não mostrados); sua estrutura global foi classificada no banco de dados SUPERFAMILY com o enovelamento "HAD-like" [56783] e Superfamília "HAD-like" [56784] (Tabela S2).

Os membros desta superfamília catalisam uma reação de substituição nucleofílica no fósforo ou centros de carbonos, usando um resíduo de ácido aspártico, o qual atua como nucleófilo e cliva ligações covalentes de substratos fosforilados (Jeremy, 2001, Collet et al., 1998). Na proteína homóloga fosfonoacetaldeído hidrolase de *Bacillus cereus*, a plataforma do sítio ativo é formada por 4 segmentos de peptídeos que fazem parte do domínio central, estas alças possuem sequências conservadas que geralmente são usadas para identificar os representantes da família HAD (G. Zhang et al., 2004). Cada uma destas alças contribui para um ou mais grupos catalíticos. A primeira alça posiciona o nucleófilo

ASP que é usado para mediar a transferência do grupo fosforil. A segunda alça posiciona um doador da ligação de hidrogênio (Thr ou Ser) para possibilitar a ligação ao substrato (como UMP, dUMP e dTMP). A terceira alça contribui com um resíduo positivamente carregado (Arg ou Lys) para orientar o eletrófilo ou nucleófilo (G. Zhang et al., 2004)

Neste trabalho, nós obtivemos o modelo tridimensional da proteína YjjG de *E. coli* através de modelagem comparativa com a proteína YfnB de *Bacillus subtilis subsp. subtilis* str. 168 e identificamos os prováveis resíduos do sítio ativo (Figura 22). Verificamos que YjjG possui os mesmos resíduos (Asp e Thr) envolvidos na catálise encontrados em outras fosfotransferases (como na fosfoglicomutase, fosfoserina fosfatase e P-typeATPase), assim como membros da família ácido 2-haloalcanóico dehalogenase (HAD).

Através do alinhamento com a sequência Fosfonoacetaldeído (PDB: 1FEZ) de *Bacillus cereus* e informações do UNIPROTKB nós identificamos o resíduo do sítio ativo (<http://www.uniprot.org/uniprot/P0A8Y1>). De acordo com estas informações, o Asp9 coordena o  $Mg^{2+}$  e serve de catálise nucleofílica, outros resíduos que participam deste processo são Asp9, Thr16, Asp176, Asp181, Lys150 (numeração em *E. coli*). O resíduo Asp9 é mostrado na Figura 22.

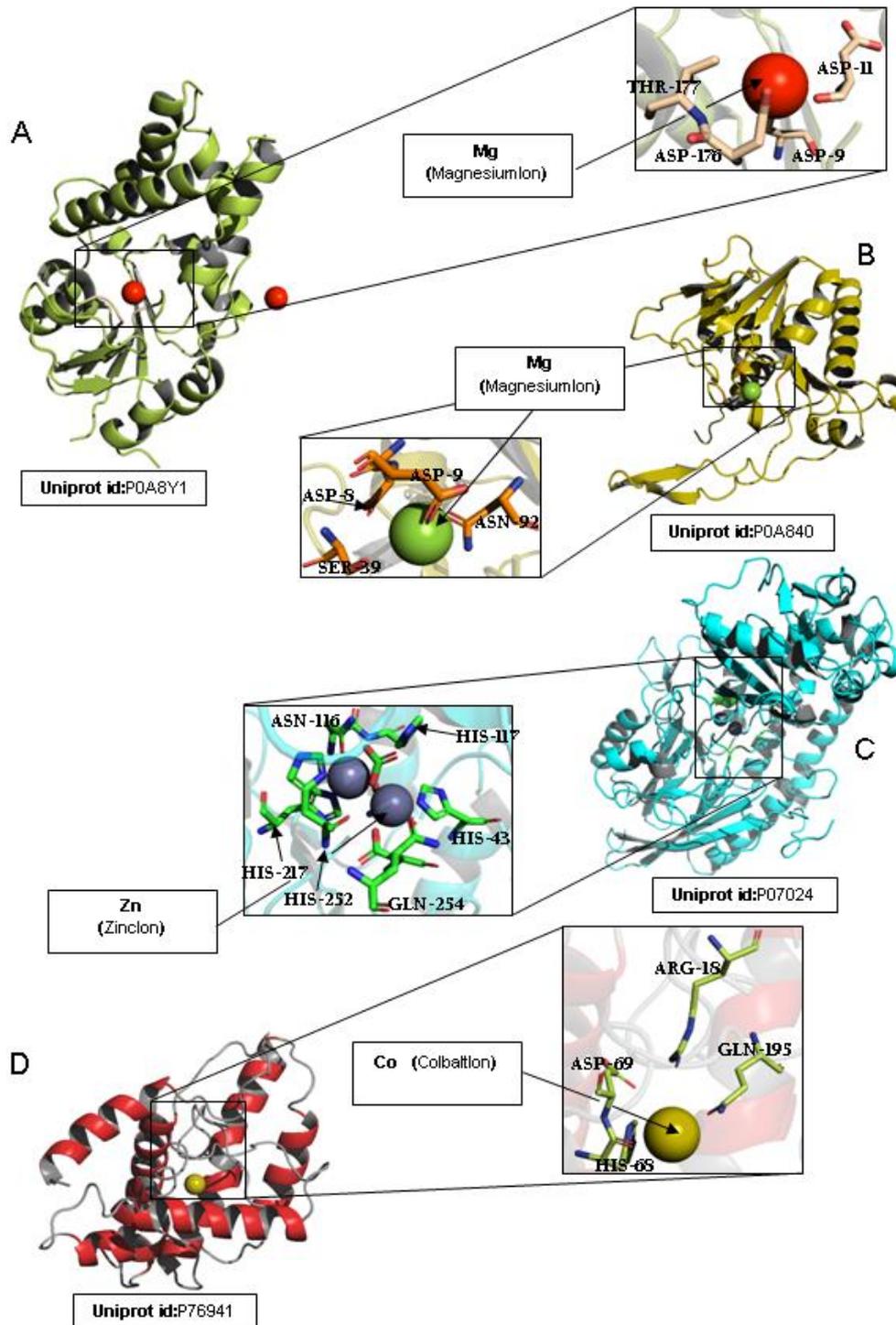


Figura 22: Modelo 3D de SurE, UshA, YfbR e YjjG. (A) Modelo obtido para a estrutura monomérica de YjjG e o resíduo Asp9 coordena o  $Mg^{2+}$  e serve para a catálise nucleofílica. B) O modelo obtido do homotetrâmero SurE, e alguns resíduos do sítio ativo, descritos por Pappachan, *et al.*, 2008. C) O homotetrâmero UshA (P07024) e a região onde está localizada a díade catalítica (Asp120 e His117). D) O homotetrâmero yfbR (P76491) e a região próxima aos resíduos do sítio ativo Glu72 ou Asp77 (Zimmerman *et al.*, 2008). A região da figura amplificada mostra detalhes do sítio ativo em uma distância de 5 Å (angstrom) do ligante, figuras foram geradas com o programa MAESTRO. PDB [ID]: 3TY2 (molde para SurE), 1USH (UshA), YfbR (2PAQ), 3ED5 (molde para YjjG).

A proteína 5'/3'-nucleotidase SurE é uma proteína multifuncional que possui especificidade contra vários substratos, podendo defosforilar vários ribo- e deoxiribonucleosídeos 5'-monofosfatos e ribonucleosídeos 3'-monofosfatos (3'-AMP, 5'-dGMP, 5'-GMP, 3'-CMP e 5'-AMP) com alta afinidade para 3'-AMP; também hidrolisa polifosfatos (atividade de exopolifosfatase) com a preferência de substratos de cadeias de curto comprimento (Proudfoot et al., 2004). A família SurE é um grupo de proteínas conservadas encontrada em nas Eubacteria, nas Archaea e nos Eucariotos.

Obtivemos o modelo 3D de SurE (Figura 22B) através de modelagem comparativa com o molde SurE de *Coxiella burnetii* e identificamos os resíduos do sítio ativo (Figura 22). Estes resíduos os mesmos presentes na sequência homóloga SurE de *S. typhimurium* (Pappachan et al., 2008), estas proteínas possuem 91.7% de identidade (dados não mostrados).

O sítio ativo de SurE de *S. typhimurium* é encontrado na interface dos dois monômeros é uma região parcialmente ácida devido a presença dos resíduos Asp8 e Asp9, o íon de  $Mg^{2+}$  é coordenada por estes dois resíduos (Pappachan et al., 2008). A estrutura de SurE possui seis regiões  $\alpha$ -hélice e entre 11 a 13 regiões de folhas- $\beta$  preditas (dados não mostrados); sua estrutura global foi classificada no banco de dados SCOP com o enovelamento SurE-like [64166] e superfamília SurE-like [64167].

Pappachan e colaboradores (2008) propuseram um mecanismo para SurE de *S. typhimurium* baseado em um ataque nucleofílico similar ao encontrado em TA017, uma fosfoglicolato fosfatase membro da superfamília HAD. Nesta transformação química mediada por SurE, possivelmente o resíduo Asp8 está envolvido no ataque nucleofílico, a interação do fosfato com  $Mn^{2+}$  no carbono C2-SurE (Asp8) pode ajudar a polarizar a ligação fosfato para ataque nucleofílico e o metal ( $Mn^{2+}$ ) pode também orientar o nucleófilo Asp8, então, este pode atacar o átomo de fósforo do monofosfato. O próximo passo requer uma molécula de água ativada para produzir um  $OH^-$ , pelo qual irá hidrolisar intermediário Asp8 fosfato para liberar o fosfato e regenerar a enzima ativa (Pappachan et al., 2008).

A forma análoga UshA é uma enzima multifuncional dependente de zinco. Esta enzima possui os resíduos de 1-25 que constituem uma sequência que sinaliza

para exportação para o espaço periplasmático, onde pode catalisar as seguintes reações: (1)  $\text{UDP-sugar} + \text{H}_2\text{O} = \text{UMP} + \text{alfa-D-aldose 1-fosphate}$ ; (2)  $5\text{'-ribonucleotideo} + \text{H}_2\text{O} = \text{a ribonucleosídeo} + \text{fosfato}$ . Através de análises no banco de dados SCOP classificamos UshA com o enovelamento “Metallo-dependent phosphatases [56299]” e “5'-nucleotidase (syn. UDP-sugar hydrolase), C-terminal domain [55815]”.

As regiões importantes para a catálise na estrutura de UshA são uma díade catalítica de His117 e Asp120 (Figura 22C), localizadas na região N-terminal, e o centro dimetal com seus resíduos ligantes e água potencialmente ligada, localizados na interface entre os dois domínios (ligados por uma  $\alpha$ -hélice de 19 resíduos) (Knöfel & Sträter, 2001, Knöfel & Sträter, 1999). Os resíduos Asp41, His43, Gln254, Asp84 estão ligados à primeira molécula de metal são e os resíduos Asn116, His217 e His252 estão ligados à segunda molécula de metal, e Arg375, Arg379, Arg419 contribuem para o potencial positivo distinto no sítio ativo, estabilizando um estado de transição (Knöfel e Sträter, 2001, Knöfel e Sträter, 1999).

A enzima UshA catalisa a clivagem hidrolítica de 5'-mononucleotídeos para nucleosídeos e fosfato. O mecanismo de transformação química proposto é um ataque nucleofílico de duas moléculas de água ou hidroxila ligadas ao átomo de  $\text{Zn}^{2+}$  que podem atacar o átomo eletrofílico de fósforo, mas várias propostas de mecanismos têm sido feitas a para hidrólise da ligação éster-fosfato mediado por Ser/ThrPP se fosfatases ácidas (PAPs) (Knöfel e Sträter, 1999).

A forma análoga YfbR (Figura 22D) é uma 5'-nucleotidase que pertence à superfamília “domínio HD”; esta enzima é específica para 2'-deoxiribonucleotídeo-5'-monofosfatos (dNMPs). A estrutura de YfbR foi classificada no banco de dados SCOP com o enovelamento “HD-domain/PDEase-like [109603]” e Superfamília “HD-domain/PDEase-like [109604]”. O mecanismo de transformação química mediada por YfbR ainda não é conclusivo. Acredita-se que a enzima atue sobre o substrato através de um mecanismo de substituição nucleofílica que envolve o resíduo Glu72 ou Asp77, o qual pode doar um próton mediado por uma ligação com água ou um ataque nucleofílico de um íon hidroxila ligada ao metal que cliva a ligação fosfodiéster do substrato liberando o fósforo; de outra forma, o ataque nucleofílico pode ser também mediado pelo oxigênio de Asp137 não coordenado ao metal, o

qual pode formar um intermediário fosfoaspartil e no passo seguinte o fosfato pode ser liberado por uma segunda reação de substituição com uma molécula de água ou uma molécula de água ligada a um resíduo ácido (ácido glutâmico e ácido aspártico), mecanismo similar ao da proteína UshA (Zimmerman *et al.*, 2008). Os resíduos importantes para atividade catalítica de YfbR são Glu72 ou Asp77, Asp137, His33, His68, Asp69 e Asp137 (resíduos de ligação a  $\text{Co}^{2+}$ ), Arg18, Glu72 e Asp77 (Zimmerman *et al.*, 2008).

Concluindo, nossas análises sugerem que, apesar de sua origem evolutiva distinta, resultando em uma estrutura terciária também distinta, houve não somente convergência estrutural como também mecanística entre as formas análogas SurE, UshA, YfbR e YjjG.

#### **4.7 Lisozima (EC 3.2.1.17): *rrrD*, *yefX*, *chiA*, endolisina**

As lisozimas são enzimas que os peptidoglicanos da parede celular de bactérias através da hidrólise de ligações glicosídicas  $\beta$ -1,4 entre os resíduos do carbono C-1 de N-ácido N-acetilmurâmico (MurNAc) e o carbono C-4 do N-acetilglicosamina (Davis & Weiser, 2011). Para a atividade enzimática lisozima identificamos 4 distintas formas estruturais, RrrD, YefX ChiA e endolisina (lisozima\_1, \_2, \_3 e \_4). O gene *rrrD* é encontrado em um maior número de cópias em algumas linhagens enterohemorrágicas. (Figura 23).

### Lisozima (EC3.2.1.17)

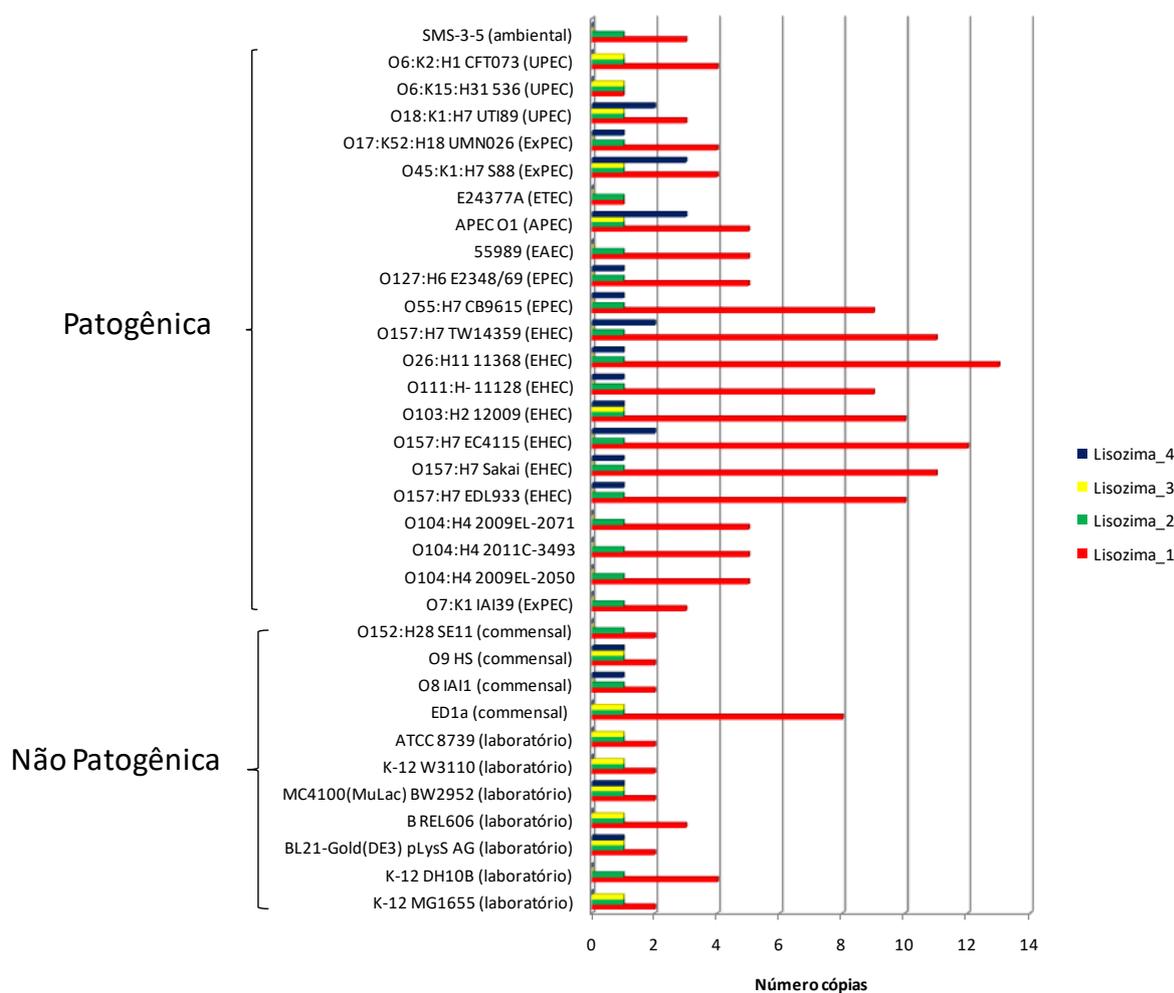


Figura 23: Número de cópias de *rrrD* (lisozima\_1), *yefX* (lisozima\_2), *chiA* (lisozima\_3) e endolisina (lisozima\_4) nas 33 linhagens estudadas. EHEC: Enterohemorrágica, EPEC: Enteropatogênica, ExPEC: Estraintestinal, UPEC: Uropatogênica.

O gene *rrrD* (Lisozima\_1) de *E. coli* K12 é uma endolisina codificada pelo profago DLP12, que como visto acima desempenha um papel importante na formação de biofilmes (Toba *et al.*, 2011b), pois foi verificado que a deleção destes profagos impede a formação de biofilmes (Wang *et al.*, 2010). As lisozimas e holinas presentes em bacteriófagos atuam na lise da parede celular da bactéria infectada, dissolvendo a parede celular rica em peptidoglicanos (Srividhya & Rishnaswamy,

2007). A morte celular em bactérias é importante para formação e desenvolvimento de biofilme nas micro-colônias; após a morte celular o DNA liberado participa na adesão e estabilidade do biofilme, a holina está envolvida no controle deste processo (Bayles, 2007), tal proteína também é codificada dentro do profago DLP12 (Krishnaswamy, 2012).

Foram identificados três casos de analogia funcional no genoma de *E. coli* K12: RrrD (primeira forma estrutural), YefX (segunda forma estrutural) e ChiA (terceira forma estrutural), a quarta forma estrutural (anotada como endolisina) não foi encontrada nesta linhagem. Por outro lado, tal forma estrutural está presente em várias linhagens patogênicas e apenas em algumas das linhagens não patogênicas (Figura 23). A enzima ChiA está presente em linhagens comensais, *Extraintestinais* (EXPEC), Uropatogênica (UPEC) e APEC01 (Figura 23).

A proteína codificada por *yefX* está anotada no UNIPROTKB como uma proteína não caracterizada, mas no banco de dados CDD do NCBI este gene é caracterizado por conter um domínio catalítico da família glicosilhidrolase que é similar para lisozimas CH-type (*Chalaropsis-type*) da família GH25 de endolisinas. Em fagos, as proteínas que degradam a parede celular são denominadas genericamente de endolisinas (Young, 1992), alguns dos genes identificados estão anotados com este termo no NCBI (dados não mostrados). O gene *chiA* codifica uma enzima bifuncional com atividade de lisozima e quitinase (EC 3.2.1.14) (Francetic *et al.*, 2000). Esta lisozima é secretada no meio extracelular (Francetic *et al.*, 2000)

Conforme mostrado na Figura 23, encontramos um aumento do número de cópias da forma estrutural lisozima\_1 (RrrD) em linhagens de *E. coli* patogênicas. A linhagem O157: H7 Sakai (EHEC), por exemplo, possui 11 cópias de *rrrD* (Figura 24) contidas em profagos no genoma.

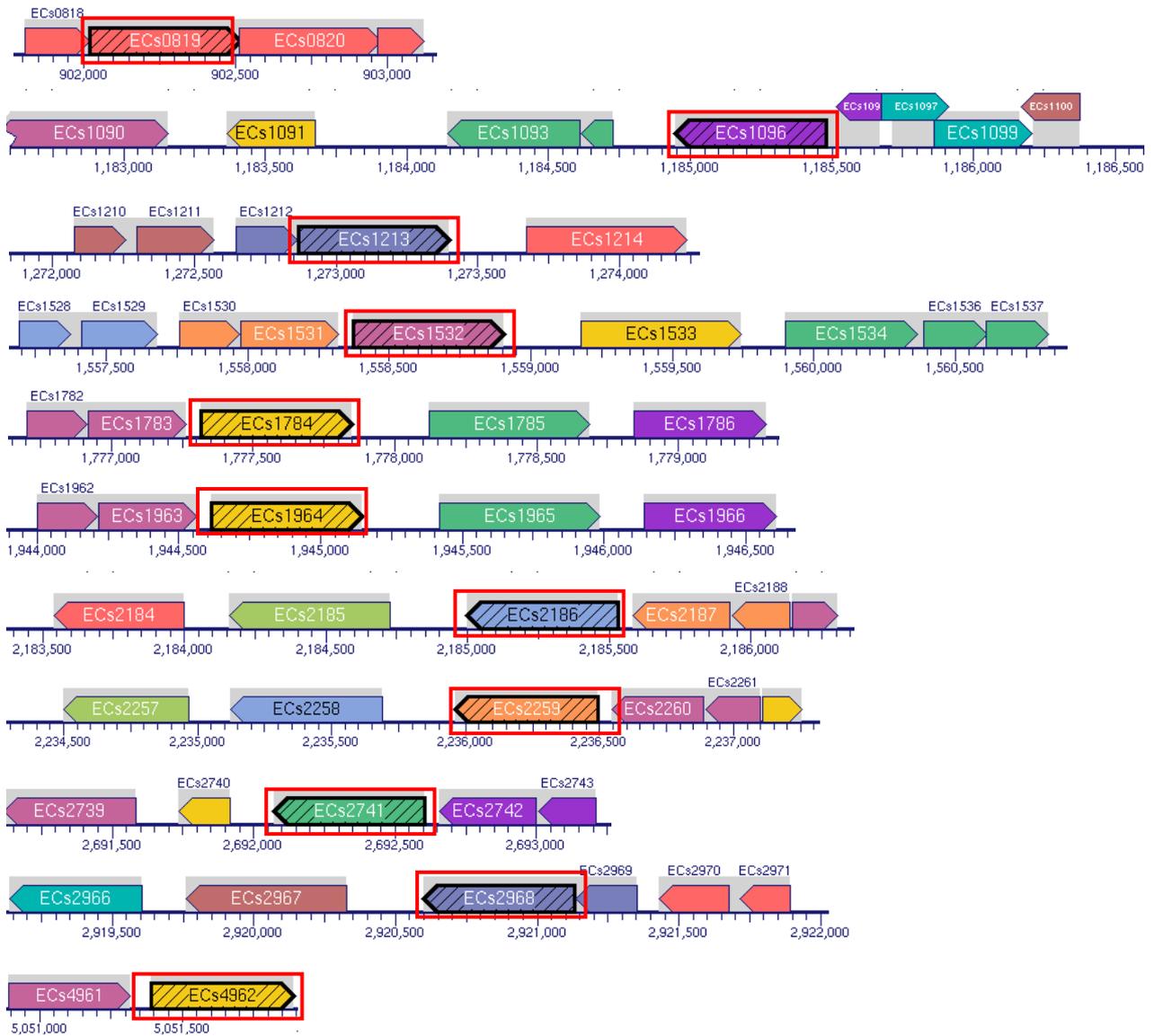


Figura 24: Contexto genômico de lisozima\_1 (*rrrD*) em *E. coli* O157:H7 Sakai (EHEC). Cada linha é mostrado um lócus no genoma (1-11) que codificam *rrrD*, cada cópia está assinalada por um retângulo vermelho. A legenda completa, com a descrição de todos os genes, encontra-se nos anexos (Figuras S2).

Na linhagem K12, a super-expressão do gene *rrrD* (lisozima\_1) resulta em lise da célula (Srividhya & Rishnaswamy, 2007). O gene *rrrD* é codificado dentro do profago DLP12, o mutante de *E. coli* para este gene não é capaz de sintetizar biofilme e isto pode ser associado a um efeito indireto devido a alterações na parede de peptidoglicano (Toba *et al.*, 2011).

Nós observamos que *upstream* ao gene *rrrD* está o gene *holin*, também associado à formação de biofilmes (Srividhya & Rishnaswamy, 2007), esta proteína atua como um relógio molecular em um processo que resulta em morte celular programada em bactérias (Bayles, 2007, Wang *et al.*, 2000). Este gene parece estar associado com toxicidade em *Clostridium difficile* e participa do transporte de proteínas tóxicas como TcdA e TcdB (Tan, Wee, & Song, 2001)

A atividade de profagos em *E. coli* têm sido diretamente associada com um aumento da taxa de crescimento em várias condições. Recentemente, foi demonstrado que a deleção dos profagos CP4-6, DLP12, e14, rac, Qin, CP4-44, CPS-53, CPZ-55 e CP4-57 resultaram em uma diminuição do crescimento de *E. coli* durante o tratamento com os antibióticos da classe quinolona: ácido oxolinico, ácido nalidico, segunda geração de enoxacina, iomfloxacina e ofloxacina (Wang *et al.*, 2010) e também aboliu completamente a formação de biofilmes.

Com base nesses dados, especulamos que o aumento no número de cópias de *rrrD* pode estar associado aos mecanismos de morte celular programada, que resulta na formação de biofilmes. Como visto, tal fenótipo pode contribuir para um aumento de *fitness* em diversas situações, inclusive processos relacionados à patogenicidade.

#### **4.8 Arginina decarboxilase (EC 4.1.1.19): SpeA e AdiA**

Na transformação química catalisada pelos análogos SpeA e AdiA ocorre a decarboxilação de uma molécula de arginina produzindo uma molécula de agmatina, consumindo um próton no processo, a reação é mostrada como: L-arginine $\rightleftharpoons$ agmatine + CO(2). As formas estruturais AdiA e SpeA estão contidas em apenas um número de cópias em todas as linhagens estudadas (Tabela S3), AdiA é altamente expressa em pH ácido (2,5) e esta enzima, juntamente com a proteína AdiC, compõe um sistema que protege a célula bacteriana durante condições de pH ácido com valores extracelulares menores que 2,5; neste processo, AdiC transporta agmatina para fora da célula em troca de arginina, prevenindo a acumulação de prótons no interior da célula (Andréll *et al.*, 2009).

A forma enzimática SpeA participa na biossíntese de poliaminas (putrescina, espermidina e espermina); a hidrólise de arginina catalisada por agmatinase resulta em putrescina e uréia (Buch & Boyle, 1985). As poliaminas estão envolvidas em vários processos celulares nas células vivas, tais como crescimento celular, diferenciação, replicação e reparo de DNA, síntese e processamento de RNA, metabolismo e síntese de proteínas (Forouhar *et al.*, 2010). Na bactéria *Yersinia pestis* os genes *speA* (arginina decarboxilase) e *speC* (ornitina decarboxilase) estão envolvidos na formação de biofilmes (Patel *et al.*, 2006); bactérias mutantes para estes genes perdem a habilidade de produzir biofilmes *in vitro*. A atividade enzimática de SpeA é induzida na presença de arginina em condições de pH neutro e meio mínimo (Morris, & Morris, 1973). As enzimas SpeA e AdiA estão localizados em distintos compartimentos celulares e possuem distintos valores de pH ótimos para sua atividade catalítica (Tabela 4), sendo portanto recrutadas e utilizadas em distintas condições fisiológicas.

Tabela 4: Parâmetros bioquímicos das enzimas análogas adiA e speA

<b>Enzima</b>	<b>Localização</b>	<b>pH</b>
AdiA	citoplasmática [PMID:20472543]	5.2 [PMID: 4870599]
SpeA	espaçoperiplásmico [PMID: 2198270]	8.4 [PMID: 4571773]

Apesar de parcial, esta análise de adiAe speA evidencia um distinto papel biológico, onde especulamos sobre uma possível participação de speA na formação de biofilmes, indicando que esta atividade também deve ser examinada em maiores detalhes.

## 5 DISCUSSÃO

As NISEs são enzimas que catalisam a mesma transformação química (função bioquímica) e que não compartilham um ancestral em comum, ou seja, são enzimas com origens evolutivas independentes originadas por processos de convergência evolutiva. Trabalhos anteriores objetivaram a identificação de casos de analogia funcional em uma série de genomas (Galperin, *et al* 1998, Omelchenko *et al.*; Galperin, *et al.*, 2010). Neste trabalho, com o objetivo de analisar o significado biológico da manutenção de NISEs em um mesmo genoma, escolhemos identificar NISEs intragenômicas em *E. coli*, por ser um dos organismos-modelo mais estudados e pela ampla disponibilidade de informações em diferentes bancos de dados biológicos.

Dentre os 33 casos de analogia identificados, através da compilação de dados da literatura sobre estudos de mutantes em *E. coli* verificamos que 10 atividades enzimáticas possuem ao menos uma forma estrutural que é essencial em condições de meio mínimo (Tabela S5), isto indica que este genes não apresentam o mesmo papel biológico, pois este termo é descrito para casos onde uma mutação ou defeito em um gene irá ter um pequeno efeito no *fitness* do organismo, pois tal função pode ser substituída por outro gene (Kanaoka *et al.*, 2005).

O trabalho realizado Gerdes e colaboradores (Gerdes *et al.*, 2003), mostraram que PpiB e SlyD (Peptidil prolil isomerase, EC 2.1.2.2) são essenciais durante condições de crescimento em meio mínimo, mas resultado similar não foi observado em outros estudos (Tabela S5), estas enzimas apresentam distintos papéis biológicos, SlyD está envolvida na inserção de Ni<sub>2+</sub> (Níquel) durante maturação das hidrogenases (Zhang *et al.*, 2005), forma um complexo com HypB e estabelece comunicação entre hidrólise de GTP e liberação de Ni<sup>2+</sup> (Kaluarachchi *et al.*, 2011).

A análise do fenótipo de crescimento de mutantes em resposta a uma perturbação (por exemplo, tratamentos químicos) ou durante uma condição fisiológica tem sido uma maneira de indentificar genes com redundância funcional (Cutler & Mccourt, 2005), pois na ausência de um gene, sua função pode ser complementada por um gene com função similar (Laeremans *et al.*, 1997).

A redundância funcional pode ser identificada pela comparação dos perfis de expressão gênica; utilizamos este parâmetro porque é um indicativo da resposta transcricional e das redes de regulação envolvidas em determinado fenótipo (Hughes *et al.*, 2000).

Algumas NISEs quando mutanteadas possuem taxas negativas de *fitness* enquanto seu análogo funcional possui uma taxa de *fitness* positiva, mas tal situação varia de acordo com as condições de tratamentos no meio de cultivo (Tabela S4), indicando que o sistema metabólico é altamente flexível. Quando tratado com Paraquat, o mutante *sodA* apresenta uma taxa de *fitness* negativa (Tabela S4). Isto indica que na ausência de *sodA* a bactéria não consegue responder aos efeitos deletérios de Paraquat, mesmo possuindo um análogo funcional para a mesma atividade enzimática. Gort e colaboradores mostraram que mutantes *sodC* não foram mais sensíveis a presença de Paraquat do que a *E. coli* do tipo selvagem (Gort *et al.*, 2002). Portanto os conjuntos de dados compilados para *sodC* e *sodA* mostram que estas enzimas não são casos de redundância funcional.

Além de *sodA* e *sodC*, os mutantes PurN e PurT (Fosforibosil glicinamida formiltransferase, EC 2.1.2.2), PldB e TesA (Lisofosfolipase, EC 3.1.1.5), PgpB e PgpA (Fosfatidilglicerofosfatase, EC 3.1.3.27), AdiA e SpeA (Arginina decarboxilase, EC 4.1.1.19), PpiA e SurA (Peptidil prolil isomerase, EC 2.1.2.2) possuem distintos perfis de *fitness* durante o tratamento com compostos que causam danos ou estresses na bactéria. Os dados de expressão indicam que tais NISEs possuem distintos padrões de expressão e co-expressão no genoma; isto sugere que o papel destas enzimas é diferente e que para cada forma enzimática existe uma via de regulação transcricional específica envolvida. Neste contexto, a identificação de redes de transcrição gênica pode contribuir no aprofundamento deste trabalho.

Com relação à localização de NISEs, algumas estão alocadas em distintos compartimentos (Tabela S8), e tal variável pode influenciar no papel biológico destas enzimas; por exemplo, SodA está localizada no citoplasma e é altamente efetiva em proteger as enzimas do citoplasma contra espécies reativas de oxigênio (Hopkin, Papazian, & Steinmane, 1992); SodC, por outro lado, é localizada no periplasma de bactérias gram-negativas e tem papel biológico na proteção contra as espécies

reativas de oxigênio produzidas pelas células hospedeiras (Wilks *et al.*, 1998, Sansone *et al.*, 2002, Pacello *et al.*, 2008, McCrean *et al.*, 2010).

As redes de interação permitem uma visão holística da interação de cada proteína com outros respectivos componentes na célula, pois cada proteína pode formar rede de associações formando um ou mais grupos funcionais. Analisamos a classificação funcional das redes de interação das NISEs, e observamos que tais enzimas podem formar diferentes grupos funcionais, como por exemplo, a rede interação AdhP (Álcool desidrogenase, EC 1.1.1.1) foi classificada em 9 grupos funcionais, enquanto AdHE foi classificada em 22 grupos funcionais (Figura S3); tal comparação pode fornecer uma evidência dos distintos papéis biológicos empregados por NISEs. Porém, para uma análise mais acurada dos resultados obtidos é necessária uma revisão minuciosa da literatura.

Os estudos de pan-genoma têm como objetivo descrever todo o complemento de genes de um organismo, variações no número de cópias gênicas, entretanto, geralmente não são mensuradas. (Laing *et al.*, 2010, Rasko *et al.*, 2008). Dentre as abordagens computacionais utilizadas neste trabalho, a análise do número de cópias das diferentes formas estruturais das atividades enzimáticas analisadas possibilitou a identificação de amplificação gênica em algumas das NISEs. Interessantemente, observamos uma possível associação entre o número de cópias de algumas destas formas com diversos mecanismos de patogenicidade em *E. coli*. Estas enzimas participam em processos biológicos ímpares, como por exemplo na resistência a antibióticos (*folA*: Di-hidrofolato redutase, EC 1.5.1.3), morte celular programada e formação de biofilmes (*rrrD*: lisozima, EC 3.2.1.17), recombinação genética (*rusA*: *Crossover junction endodeoxiribonuclease*, EC 3.1.22.4) e subversão da resposta imune do hospedeiro (*ushA*: 5'-nucleotidase), detoxificação de espécies reativas de oxigênio (*sodC*: Superóxido dismutase, EC 1.15.1.1).

Isto sugere que amplificações no metabolismo central podem ter um papel fundamental na adaptação e sobrevivência de *E. coli* em seus hospedeiros. De forma geral, a maioria dos estudos tem associado estes processos biológicos às variações presentes no genoma acessório, através da perda e aquisição de genes; nós mostramos que estes processos também podem ser encontrados no genoma

núcleo através de amplificação gênica, em alguns casos possivelmente devido à integração de profagos no genoma ou a eventos de transposição.

Dentre as enzimas que identificamos no metabolismo central e genoma núcleo, a amplificação de *sodC* já havia sido descrita em *E. coli* O: 157, por D'Orazio e colaboradores (D'Orazio *et al.*, 2008) que propuseram que tal fenômeno pode estar associado à virulência, pois as duas cópias de *sodC* oriundas de profagos possuem diferentes propriedades físicas químicas, tais como eficiência catalítica, afinidade ao metal, resistência a protease e estabilidade, o que torna estas enzimas mais adequadas para resistir às condições encontradas por estas bactérias nos hospedeiros (D'Orazio *et al.*, 2008).

A enzima F<sub>olA</sub> é uma enzima chave e essencial para o crescimento de *E. coli* e tem sido largamente estudada como alvo para drogas. Recentes estudos demonstraram que a amplificação artificial (em laboratório) deste gene através de superexpressão é um determinante para a resistência ao antibiótico trimetoprim (Soo *et al.*, 2011). Em nosso estudo observamos que linhagens O104: H4 causadoras de surtos endêmicos contêm uma cópia adicional deste gene, este parece ser o primeiro trabalho que descreve amplificação de *folA* em isolados naturais de *E. coli*. Entretanto, neste caso particular isso deve ser confirmado com um número maior de linhagens, pois esta amplificação foi verificada em apenas três linhagens muito próximas.

A amplificação gênica é um processo que pode contribuir para uma rápida adaptação pelo aumento simultâneo da expressão de genes que estão agrupados no cromossomo. A amplificação gênica tem sido detectada como um fenômeno transitório em culturas de bactérias (Brochet *et al.*, 2008). Foi observado inicialmente que muitas amplificações são instáveis, e tipicamente desaparecem na ausência de seleção após somente umas poucas gerações de crescimento (Brochet *et al.*, 2008); mas quando sujeita a uma forte pressão seletiva, esta amplificação pode estar presente em 100% das células em uma população (Andersson & Hughes, 2009).

O Sistema SOS é um estado de alta atividade de reparo de DNA em células de bactérias expostas a agentes mutagênicos; esta resposta tem uma forte ligação com processos adaptativos e está envolvida em diversos fenótipos significantes, tais como a disseminação de fatores de virulência (Guerin *et al.*, 2009), formação de

biofilmes, células de persistência e resistência a drogas (Bernier et al., 2013). Em um paciente hospitalizado, a emergência de um isolado de *P. aeruginosa* resistente a antibióticos (cefalosporinas, ceftazimida) foi associada com a recombinação de um integron induzida pela resposta SOS (Hocquet et al., 2012).

Circuitos adicionais de resposta SOS regulados por LexA foram descritos presentes em regiões *upstream* a integrons cromossômicos e móveis; a regulação destas integrases é dependente da resposta SOS (Guerin et al., 2009). Nós sugerimos que a amplificação de *lexA-rusA* (Figuras S2), localizadas dentro de profagos no genoma de algumas linhagens patogênicas de *E. coli* pode fornecer um circuito adicional de resposta ao sistema SOS, podendo ser um determinante de resistência presente em linhagens enterohemorrágicas, aumentando a taxa de recombinação genética nestas bactérias patogênicas.

As linhagens EPEC não produzem toxinas e o modo pelo qual tais bactérias causam diarreia não é conhecido, Crane e colaboradores (Crane et al., 2002), propuseram que, durante a infecção por linhagens EPEC, a atividade 5'-nucleotidase (*ushA*) da célula hospedeira ativa a liberação de ATP que funciona como um mediador que causa a diarreia. Em nosso trabalho, verificamos uma amplificação do número de cópias de *ushA* em linhagens patogênicas de *E. coli*, mostrando uma possível associação com a patogenicidade.

Conforme mostrado, ocorreu uma expansão de lisozimas (RrqD) em linhagens patogênicas de *E. coli*. A literatura indica um possível papel biológico desta forma estrutural com mecanismos de lise celular, formação de biofilme; aderência celular e necrose da célula hospedeira (Höring, et al 2012)

Finalizando, as enzimas análogas intragenômicas, apesar do seu pequeno número em relação ao total de enzimas em *E. coli*, parecem realmente desempenhar papéis distintos, não sendo casos de redundância funcional. Interessantemente, várias parecem estar envolvidas com diversos mecanismos relacionados à patogenicidade, de uma forma geral, mas poucos dados sobre essas associações foram encontrados na literatura. Supomos que isso possa estar relacionado ao fato da maioria dos estudos visando a identificação de genes relacionados à patogenicidade procurar por genes específicos de linhagens patogênicas ou virulentas (em outras palavras, presentes no genoma acessório),

relegando genes compartilhados (presentes no genoma núcleo) a um segundo plano. A análise de características simples, como o número de cópias de cada forma estrutural no genoma, foi sugestiva de uma possível associação entre essas atividades enzimáticas, e mais particularmente de uma forma estrutural específica, com mecanismos de patogenicidade. Além disso, tentamos, com base nos dados gerados e na análise da literatura, propor um modelo que explicasse o possível papel biológico de algumas NISEs.

## 6 CONCLUSÕES

As enzimas análogas intragenômicas não representam casos de redundância funcional. A maioria dos dados obtidos aponta para um distinto papel biológico para cada forma estrutural alternativa das atividades enzimáticas encontradas neste grupo de enzimas.

As enzimas análogas intragenômicas aparentam desempenhar importantes funções no metabolismo de *Escherichia coli*, com a maioria das mesmas sendo encontradas no genoma núcleo e/ou participando do seu metabolismo central.

Algumas NISE aparentam estar associadas com diversos mecanismos de patogenicidade em *E. coli*. Algumas atividades enzimáticas apresentam uma possível ligação entre o número de cópias de uma de suas formas estruturais particulares com algum destes mecanismos, como a formação de biofilmes, detoxificação de drogas, morte celular programada e formação de células de persistência.

## 7 REFERÊNCIAS

- Ahmed, S. a, Awosika, J., Baldwin, C., Bishop-Lilly, K. a, Biswas, B., Broomall, S., ... Gibbons, H. S. (2012). Genomic comparison of *Escherichia coli* O104:H4 isolates from 2009 and 2011 reveals plasmid, and prophage heterogeneity, including shiga toxin encoding phage stx2. *PloS One*, 7(11), e48228. doi:10.1371/journal.pone.0048228
- Ahrweiler, P. M., & Frieden, C. (1988). Construction of a fol mutant strain of *Escherichia coli* for use in dihydrofolate Construction of a fol Mutant Strain of *Escherichia coli* for Use in Dihydrofolate Reductase Mutagenesis Experiments, 170(7).
- Akashi, H. (2003). Metabolic economics and microbial proteome evolution, 19, 2003. doi:10.1093/bioinformatics/btg1053
- Akashi, H., & Gojobori, T. (2002). Metabolic efficiency and amino acid composition in the proteomes of *Escherichia coli* and *Bacillus subtilis*. *Proceedings of the National Academy of Sciences of the United States of America*, 99(6), 3695–700. doi:10.1073/pnas.062526999
- Altschul, S. F., Madden, T. L., Schäffer, a a, Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, 25(17), 3389–402.
- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, 25(17), 3389–3402.
- Andersen, C. L., Holland, I. B., & Jacq, a. (2006). Verapamil, a Ca<sup>2+</sup> channel inhibitor acts as a local anesthetic and induces the sigma E dependent extra-cytoplasmic stress response in *E. coli*. *Biochimica et Biophysica Acta*, 1758(10), 1587–95. doi:10.1016/j.bbamem.2006.05.022
- Andersson, D. I., & Hughes, D. (2009). Gene amplification and adaptive evolution in bacteria. *Annual Review of Genetics*, 43, 167–95. doi:10.1146/annurev-genet-102108-134805
- Andréll, J., Hicks, M. G., Palmer, T., Carpenter, E. P., Iwata, S., & Maher, M. J. (2009). Crystal structure of the acid-induced arginine decarboxylase from *Escherichia coli*: reversible decamer assembly controls enzyme activity. *Biochemistry*, 48(18), 3915–27. doi:10.1021/bi900075d
- Ariyoshi, M., Vassylyev, D. G., Iwasaki, H., Nakamura, H., Shinagawa, H., & Morikawa, K. (1994). Atomic structure of the RuvC resolvase: a holliday junction-specific endonuclease from *E. coli*. *Cell*, 78(6), 1063–72.

- Askari, B. S., & Krajinovic, M. (2010). Dihydrofolate reductase gene variations in susceptibility to disease and treatment outcomes. *Current Genomics*, 11(8), 578–83. doi:10.2174/138920210793360925
- Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., ... Mori, H. (2006). Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular Systems Biology*, 2, 2006.0008. doi:10.1038/msb4100050
- Barrack, K. L., Tulloch, L. B., Burke, L.-A., Fyfe, P. K., & Hunter, W. N. (2011). Structure of recombinant Leishmania donovani pteridine reductase reveals a disordered active site. *Acta Crystallographica. Section F, Structural Biology and Crystallization Communications*, 67(Pt 1), 33–7. doi:10.1107/S174430911004724X
- Baker D, Sali A. Protein structure prediction and structural genomics. *Science*. 2001 Oct 5; 294(5540):93-6.
- Battistoni, a. (2003). Role of prokaryotic Cu,Zn superoxide dismutase in pathogenesis. *Biochemical Society Transactions*, 31(Pt 6), 1326–9. doi:10.1042/
- Bayles, K. W. (2007). The biological role of death and lysis in biofilm development, 5(September), 721–726.
- Benner, S. A. (1989). Enzyme Kinetics and Molecular Evolution, 789–806.
- Bennett, B., Langan, P., Coates, L., Mustyakimov, M., Schoenborn, B., Howell, E. E., & Dealwis, C. (2006). Neutron diffraction studies of Escherichia coli dihydrofolate reductase complexed with methotrexate. *Proceedings of the National Academy of Sciences of the United States of America*, 103(49), 18493–8. doi:10.1073/pnas.0604977103
- Bennett, R. J., Dunderdale, H. J., & West, S. C. (1993). Resolution of Holliday junctions by RuvC resolvase: cleavage specificity and DNA distortion. *Cell*, 74(6), 1021–31.
- Bernier, S. P., Lebeaux, D., DeFrancesco, A. S., Valomon, A., Soubigou, G., Coppée, J.-Y., ... Beloin, C. (2013). Starvation, together with the SOS response, mediates high biofilm-specific tolerance to the fluoroquinolone ofloxacin. *PLoS Genetics*, 9(1), e1003144. doi:10.1371/journal.pgen.1003144
- Bettelheim, K. A. (1976). The Acquisition of Escherichia coli by New-Born Babies Colonization of Normally Delivered Infants, 4(3), 174–179.
- Bielaszewska, M., Mellmann, A., Zhang, W., Köck, R., Fruth, A., Bauwens, A., ... Karch, H. (2011). Characterisation of the Escherichia coli strain associated with an outbreak of haemolytic uraemic syndrome in Germany, 2011: a microbiological study. *The Lancet. Infectious Diseases*, 11(9), 671–6. doi:10.1016/S1473-3099(11)70165-7

- Biopolymers, W. S., Vol, E., Dekket, M., York, N., Enzymol, P. C. M., Wierenga, R. K., ... Hol, W. G. J. (1985). Interaction of Pyrophosphate Moieties with  $\alpha$ -Helices in Dinucleotide Binding Proteins, (1976), 672–680.
- Bodewits, K., Raetz, C. R. H., Govan, J. R., & Campopiano, D. J. (2010). Antimicrobial activity of CHIR-090, an inhibitor of lipopolysaccharide biosynthesis, against the *Burkholderia cepacia* complex. *Antimicrobial Agents and Chemotherapy*, 54(8), 3531–3. doi:10.1128/AAC.01600-09
- Bogan, K. L. (2010). 5' Nucleotidases and their new roles in NAD<sup>+</sup> and phosphate metabolism w, 845–853. doi:10.1039/b9nj00758j
- Bolt, E. L., & Lloyd, R. G. (2002). Substrate specificity of RuvA resolvase reveals the DNA structures targeted by RuvAB and RecG in vivo. *Molecular Cell*, 10(1), 187–98.
- Boyd, E. F., Almagro-Moreno, S., & Parent, M. a. (2009). Genomic islands are dynamic, ancient integrative elements in bacterial evolution. *Trends in Microbiology*, 17(2), 47–53. doi:10.1016/j.tim.2008.11.003
- Bradley, A. S., Baharoglu, Z., Niewiarowski, A., Michel, B., & Tsaneva, I. R. (2011). Formation of a stable RuvA protein double tetramer is required for efficient branch migration in vitro and for replication fork reversal in vivo. *The Journal of Biological Chemistry*, 286(25), 22372–83. doi:10.1074/jbc.M111.233908
- Bragg, J. G., & Wagner, A. (2007). Protein carbon content evolves in response to carbon availability and may influence the fate of duplicated genes. *Proceedings. Biological Sciences / The Royal Society*, 274(1613), 1063–70. doi:10.1098/rspb.2006.0290
- Brochet, M., Couvé, E., Zouine, M., Poyart, C., & Glaser, P. (2008a). A naturally occurring gene amplification leading to sulfonamide and trimethoprim resistance in *Streptococcus agalactiae*. *Journal of Bacteriology*, 190(2), 672–80. doi:10.1128/JB.01357-07
- Brochet, M., Couvé, E., Zouine, M., Poyart, C., & Glaser, P. (2008b). A naturally occurring gene amplification leading to sulfonamide and trimethoprim resistance in *Streptococcus agalactiae*. *Journal of Bacteriology*, 190(2), 672–80. doi:10.1128/JB.01357-07
- Buch, J. K., & Boyle, S. M. (1985). Escherichia coli is synthesized as a precursor Biosynthetic Arginine Decarboxylase in Escherichia coli Is Synthesized as a Precursor and Located in the Cell Envelope, 163(2).
- Butala, M., Klose, D., Hodnik, V., Rems, A., Podlesek, Z., Klare, J. P., ... Zgur-Bertok, D. (2011). Interconversion between bound and free conformations of LexA orchestrates the bacterial SOS response. *Nucleic Acids Research*, 39(15), 6546–57. doi:10.1093/nar/gkr265

- Capriles, P. V. S. Z., Guimarães, A. C. R., Otto, T. D., Miranda, A. B., Dardenne, L. E., & Degraeve, W. M. (2010). Structural modelling and comparative analysis of homologous, analogous and specific proteins from *Trypanosoma cruzi* versus *Homo sapiens*: putative drug targets for chagas' disease treatment. *BMC Genomics*, *11*, 610. doi:10.1186/1471-2164-11-610
- Carfi, A., Pares, S., Duee, E., Galleni, M., Duez, C., Frere, J. M., & Dideberg, O. (1995). The 3-D structure of a zinc metallo-1-lactamase from *Bacillus cereus* reveals a new type of protein fold, *14*(20), 4914–4921.
- Chen, Z., Yang, H., & Pavletich, N. P. (2008). Mechanism of homologous recombination from the RecA-ssDNA/dsDNA structures. *Nature*, *453*(7194), 489–4. doi:10.1038/nature06971
- Choowongkamon, K., Theppabutr, S., Songtawee, N., Day, N. P. J., White, N. J., Woodrow, C. J., & Imwong, M. (2010). Computational analysis of binding between malarial dihydrofolate reductases and anti-folates. *Malaria Journal*, *9*, 65. doi:10.1186/1475-2875-9-65
- Clermont, O., Lescat, M., Brien, C. L. O., Gordon, D. M., Tenailon, O., & Denamur, E. (2008). Evidence for a human-specific *Escherichia coli* clone, *10*, 1000–1006. doi:10.1111/j.1462-2920.2007.01520.x
- Collet, J.-F., Stroobant, V., Pirard, M., Delpierre, G., & Van Schaftingen, E. (1998). A New Class of Phosphotransferases Phosphorylated on an Aspartate Residue in an Amino-terminal DXDX(T/V) Motif. *Journal of Biological Chemistry*, *273*(23), 14107–14112. doi:10.1074/jbc.273.23.14107
- Cox, G. M., Harrison, T. S., McDade, H. C., Taborda, C. P., Heinrich, G., Casadevall, A., & Perfect, J. R. (2003). Superoxide dismutase influences the virulence of *Cryptococcus neoformans* by affecting growth within macrophages. *Infection and Immunity*, *71*(1), 173–80. doi:10.1128/IAI.71.1.173
- Cox, M. M. (2007). Regulation of bacterial RecA protein function. *Critical Reviews in Biochemistry and Molecular Biology*, *42*(1), 41–63. doi:10.1080/10409230701260258
- Crane, J. K., Olson, R. A., Jones, H. M., Duffey, M. E., John, K., Olson, R. A., ... Duffey, M. E. (2002). Release of ATP during host cell killing by enteropathogenic *E. coli* and its role as a secretory mediator, *142*14, 74–86.
- Crossman, L. C., Chaudhuri, R. R., Beatson, S. a, Wells, T. J., Desvaux, M., Cunningham, A. F., ... Henderson, I. R. (2010). A commensal gene bad: complete genome sequence of the prototypical enterotoxigenic *Escherichia coli* strain H10407. *Journal of Bacteriology*, *192*(21), 5822–31. doi:10.1128/JB.00710-10
- Cunningham, M. L., & Fairlamb, a H. (1995). Trypanothione reductase from *Leishmania donovani*. Purification, characterisation and inhibition by trivalent antimonials. *European Journal of Biochemistry / FEBS*, *230*(2), 460–8.

- Cutler, S., & McCourt, P. (2005). Dude , Where ' s My Phenotype ? Dealing with Redundancy in Signaling Networks, *138*(June), 558–559. doi:10.1104/pp.104.900152.558
- D’Orazio, M., Scotti, R., Nicolini, L., Cervoni, L., Rotilio, G., Battistoni, A., & Gabbianelli, R. (2008). Regulatory and structural properties differentiating the chromosomal and the bacteriophage-associated Escherichia coli O157:H7 Cu, Zn superoxide dismutases. *BMC Microbiology*, *8*, 166. doi:10.1186/1471-2180-8-166
- Daley, D. O., Rapp, M., Granseth, E., Melén, K., Drew, D., & von Heijne, G. (2005). Global topology analysis of the Escherichia coli inner membrane proteome. *Science (New York, N.Y.)*, *308*(5726), 1321–3. doi:10.1126/science.1109730
- Davis, K. M., & Weiser, J. N. (2011). Modifications to the Peptidoglycan Backbone Help Bacteria To Establish Infection Modifications to the Peptidoglycan Backbone Help Bacteria To Establish Infection □, *79*(2). doi:10.1128/IAI.00651-10
- Dawson, A., Tulloch, L. B., Barrack, K. L., & Hunter, W. N. (2010). High-resolution structures of Trypanosoma brucei pteridine reductase ligand complexes inform on the placement of new molecular entities in the active site of a potential drug target. *Acta Crystallographica. Section D, Biological Crystallography*, *66*(Pt 12), 1334–40. doi:10.1107/S0907444910040886
- Dhar, M. S., Gupta, V., & Viridi, J. S. (2013). Detection, distribution and characterization of novel superoxide dismutases from Yersinia enterocolitica Biovar 1A. *PloS One*, *8*(5), e63919. doi:10.1371/journal.pone.0063919
- Dobrindt, U. (2005). (Patho-)Genomics of Escherichia coli. *International Journal of Medical Microbiology : IJMM*, *295*(6-7), 357–71. doi:10.1016/j.ijmm.2005.07.009
- Doolittle, R. F. (1994). TALKING Convergent evolution : the need to be explicit TALKING, (January), 15–18.
- Dozois, C. M., & Iii, R. C. (1998). Review article Pathogenic diversity of Escherichia coli and the emergence of ' exotic ' islands in the gene stream.
- Duval-Valentin, G., Marty-Cointin, B., & Chandler, M. (2004). Requirement of IS911 replication before integration defines a new bacterial transposition pathway. *The EMBO Journal*, *23*(19), 3897–906. doi:10.1038/sj.emboj.7600395
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–7. doi:10.1093/nar/gkh340
- Eramian, D., Shen, M., Devos, D., Melo, F., Sali, A., & Marti-renom, M. A. (2006). A composite score for predicting errors in protein structure models, 1653–1666. doi:10.1110/ps.062095806.genomes.

- Escobar-Páramo, P., Clermont, O., Blanc-Potard, A.-B., Bui, H., Le Bouguéneq, C., & Denamur, E. (2004). A specific genetic background is required for acquisition and expression of virulence factors in *Escherichia coli*. *Molecular Biology and Evolution*, 21(6), 1085–94. doi:10.1093/molbev/msh118
- Fan, J., Zhang, Y., Chuang-smith, O. N., Frank, K. L., Guenther, B. D., Kern, M., ... Herzberg, M. C. (2012). Ecto-5 9 -Nucleotidase : A Candidate Virulence Factor in *Streptococcus sanguinis* Experimental Endocarditis, 7(6), 1–10. doi:10.1371/journal.pone.0038059
- Farewell, A., Kvint, K., Nyström, T., & Nystro, T. (1998). uspB , a New  $\zeta$  S -Regulated Gene in *Escherichia coli* Which Is Required for Stationary-Phase Resistance to Ethanol These include: uspB , a New S -Regulated Gene in *Escherichia coli* Which Is Required for Stationary-Phase Resistance to Ethanol, 180(23).
- Fasani, R. a, & Savageau, M. a. (2013). Molecular mechanisms of multiple toxin-antitoxin systems are coordinated to govern the persistor phenotype. *Proceedings of the National Academy of Sciences of the United States of America*, 110(27), E2528–37. doi:10.1073/pnas.1301023110
- Fernández De Henestrosa, a R., Ogi, T., Aoyagi, S., Chafin, D., Hayes, J. J., Ohmori, H., & Woodgate, R. (2000). Identification of additional genes belonging to the LexA regulon in *Escherichia coli*. *Molecular Microbiology*, 35(6), 1560–72.
- Forouhar, F., Lew, S., Seetharaman, J., Xiao, R., Acton, T. B., Montelione, G. T., & Tong, L. (2010). Structures of bacterial biosynthetic arginine decarboxylases. *Acta Crystallographica. Section F, Structural Biology and Crystallization Communications*, 66(Pt 12), 1562–6. doi:10.1107/S1744309110040649
- Foster, J. M., Davis, P. J., Raverdy, S., Sibley, M. H., Raleigh, E. a, Kumar, S., & Carlow, C. K. S. (2010). Evolution of bacterial phosphoglycerate mutases: non-homologous isofunctional enzymes undergoing gene losses, gains and lateral transfers. *PloS One*, 5(10), e13576. doi:10.1371/journal.pone.0013576
- Francetic, O., Badaut, C., Rimsky, S., & Pugsley, a P. (2000). The ChiA (YheB) protein of *Escherichia coli* K-12 is an endochitinase whose gene is negatively controlled by the nucleoid-structuring protein H-NS. *Molecular Microbiology*, 35(6), 1506–17.
- Francetic, O., Belin, D., Badaut, C., & Pugsley, a P. (2000). Expression of the endogenous type II secretion pathway in *Escherichia coli* leads to chitinase secretion. *The EMBO Journal*, 19(24), 6697–703. doi:10.1093/emboj/19.24.6697
- Galperin, M. Y., & Koonin, E. V. (2012a). Divergence and convergence in enzyme evolution. *The Journal of Biological Chemistry*, 287(1), 21–8. doi:10.1074/jbc.R111.241976
- Galperin, M. Y., & Koonin, E. V. (2012b). Divergence and convergence in enzyme evolution. *The Journal of Biological Chemistry*, 287(1), 21–8. doi:10.1074/jbc.R111.241976

- Galperin, M. Y., Walker, D. R., & Koonin, E. V. (1998). Analogous Enzymes : Independent Inventions in Enzyme Evolution Analogous Enzymes : Independent Inventions in Enzyme Evolution, 779–790. doi:10.1101/gr.8.8.779
- García-Contreras, R., Zhang, X.-S., Kim, Y., & Wood, T. K. (2008). Protein translation and cell death: the role of rare tRNAs in biofilm formation and in activating dormant phage killer genes. *PloS One*, 3(6), e2394. doi:10.1371/journal.pone.0002394
- Gerdes, S. Y., Scholle, M. D., Campbell, J. W., Balázsi, G., Daugherty, M. D., Somera, A. L., ... Bala, G. (2003). Experimental Determination and System Level Analysis of Essential Genes in Escherichia coli MG1655 Experimental Determination and System Level Analysis of Essential Genes in Escherichia coli MG1655. doi:10.1128/JB.185.19.5673
- Gherardini, P. F., Wass, M. N., Helmer-Citterich, M., & Sternberg, M. J. E. (2007a). Convergent evolution of enzyme active sites is not a rare phenomenon. *Journal of Molecular Biology*, 372(3), 817–45. doi:10.1016/j.jmb.2007.06.017
- Gherardini, P. F., Wass, M. N., Helmer-Citterich, M., & Sternberg, M. J. E. (2007b). Convergent evolution of enzyme active sites is not a rare phenomenon. *Journal of Molecular Biology*, 372(3), 817–45. doi:10.1016/j.jmb.2007.06.017
- Giese, K. C., Michalowski, C. B., & Little, J. W. (2008). RecA-dependent cleavage of LexA dimers. *Journal of Molecular Biology*, 377(1), 148–61. doi:10.1016/j.jmb.2007.12.025
- Giladi, M., Altman-price, N., Levin, I., Levy, L., & Mevarech, M. (2003). FoIM , A New Chromosomally Encoded Dihydrofolate Reductase in Escherichia coli FoIM , A New Chromosomally Encoded Dihydrofolate Reductase in Escherichia coli. doi:10.1128/JB.185.23.7015
- Gordon, D. M., Bauer, S., & Johnson, J. R. (2002). The genetic structure of Escherichia coli populations in primary and secondary habitats. *Microbiology (Reading, England)*, 148(Pt 5), 1513–22.
- Gort, a S., Ferber, D. M., & Imlay, J. a. (1999). The regulation and role of the periplasmic copper, zinc superoxide dismutase of Escherichia coli. *Molecular Microbiology*, 32(1), 179–91.
- Gourley, D. G., Schüttelkopf, A. W., Leonard, G. A., Luba, J., Hardy, L. W., Beverley, S. M., & Hunter, W. N. (2001). letters Pteridine reductase mechanism correlates pterin metabolism with drug resistance in trypanosomatid parasites, 8(6), 521–525.
- Guerin, É., Cambray, G., Sanchez-alberola, N., Campoy, S., Erill, I., Re, S. Da, ... Mazel, D. (2009). BREVIA The SOS Response Controls Integron Recombination, 324(May), 2009.

- Hansen, S., Lewis, K., & Vulić, M. (2008). Role of global regulators and nucleotide metabolism in antibiotic tolerance in *Escherichia coli*. *Antimicrobial Agents and Chemotherapy*, *52*(8), 2718–26. doi:10.1128/AAC.00144-08
- Hays, F. a, Watson, J., & Ho, P. S. (2003). Caution! DNA crossing: crystal structures of Holliday junctions. *The Journal of Biological Chemistry*, *278*(50), 49663–6. doi:10.1074/jbc.R300033200
- Hemm, M. R., Paul, B. J., Schneider, T. D., Storz, G., & Rudd, K. E. (2008). Small membrane proteins found by comparative genomics and ribosome binding site models. *Molecular Microbiology*, *70*(6), 1487–501. doi:10.1111/j.1365-2958.2008.06495.x
- Hocquet, D., Llanes, C., Thouverez, M., Kulasekara, H. D., Bertrand, X., Plésiat, P., ... Miller, S. I. (2012). Evidence for induction of integron-based antibiotic resistance by the SOS response in a clinical setting. *PLoS Pathogens*, *8*(6), e1002778. doi:10.1371/journal.ppat.1002778
- Hopkin, K. A., Papazian, M. A., & Steinman, H. M. (1992). Functional Differences between Manganese and Iron Superoxide Dismutases in *Escherichia coli* K- IZ \*.
- Höring, S., Schütz, M., Autenrieth, I. B., & Gröbner, S. (2012). Lysozyme facilitates adherence of *Enterococcus faecium* to host cells and induction of necrotic cell death. *Microbes and Infection / Institut Pasteur*, *14*(6), 554–62. doi:10.1016/j.micinf.2012.01.005
- Howell, E. E., Foster, P. G., & Foster, L. M. (1988). Construction of a Dihydrofolate Reductase-Deficient Mutant of *Escherichia coli* by Gene Replacement, *170*(7), 3040–3045.
- Hughes, T. R., Marton, M. J., Jones, A. R., Roberts, C. J., Stoughton, R., Armour, C. D., ... Avenue, N. E. (2000). Functional Discovery via a Compendium of Expression Profiles, *102*, 109–126.
- Induces, P., Stress, O., Human, I. N., The, M., & Redox, G. (2002). PYOCYANIN INDUCES OXIDATIVE STRESS IN HUMAN ENDOTHELIAL, *33*(11), 1527–1533.
- Jomvall, H., & Persson, B. (1995). *Biochemistry* 0, *34*(18).
- Kanehisa, M., & Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, *28*(1), 27–30.
- Kargalioglu, Y., & Imlay, J. A. (1994). Importance of Anaerobic Superoxide Dismutase Synthesis in Facilitating Outgrowth of *Escherichia coli* upon Entry into an Aerobic Habitat.
- Kim, H. K., Kim, H., Schneewind, O., & Missiakas, D. (2011). Identifying protective antigens of *Staphylococcus aureus*, a pathogen that suppresses host immune responses. doi:10.1096/fj.11-187963

- Knöfel, T., & Sträter, N. (2001). Mechanism of hydrolysis of phosphate esters by the dimetal center of 5'-nucleotidase based on crystal structures. *Journal of Molecular Biology*, 309(1), 239–54. doi:10.1006/jmbi.2001.4656
- Korshunov, S. S., & Imlay, J. a. (2002). A potential role for periplasmic superoxide dismutase in blocking the penetration of external superoxide into the cytosol of Gram-negative bacteria. *Molecular Microbiology*, 43(1), 95–106.
- Kotera, M., Goto, S., & Kanehisa, M. (2014). Predictive genomic and metabolomic analysis for the standardization of enzyme data. *Perspectives in Science*, 1(1-6), 24–32. doi:10.1016/j.pisc.2014.02.003
- Kotera, M., Okuno, Y., Hattori, M., Goto, S., & Kanehisa, M. (2004a). Computational assignment of the EC numbers for genomic-scale analysis of enzymatic reactions. *Journal of the American Chemical Society*, 126(50), 16487–98. doi:10.1021/ja0466457
- Kotera, M., Okuno, Y., Hattori, M., Goto, S., & Kanehisa, M. (2004b). Computational assignment of the EC numbers for genomic-scale analysis of enzymatic reactions. *Journal of the American Chemical Society*, 126(50), 16487–98. doi:10.1021/ja0466457
- Krishnaswamy, K. V. S. S. (2012). Characterization of DLP12 Prophage Membrane Associated Protein : HolinGFP, 52(3), 464–471. doi:10.1007/s12088-012-0278-7
- Kühner, F., Costa, L. T., Bisch, P. M., Thalhammer, S., Heckl, W. M., & Gaub, H. E. (2004). LexA-DNA bond strength by single molecule force spectroscopy. *Biophysical Journal*, 87(4), 2683–90. doi:10.1529/biophysj.104.048868
- Kumar, A., Zhang, M., Zhu, L., Liao, R. P., Mutai, C., Hafsath, S., ... Wang, M.-W. (2012). High-throughput screening and sensitized bacteria identify an M. tuberculosis dihydrofolate reductase inhibitor with whole cell activity. *PloS One*, 7(6), e39961. doi:10.1371/journal.pone.0039961
- Kussell, E., Kishony, R., Balaban, N. Q., & Leibler, S. (2005). Bacterial persistence: a model of survival in changing environments. *Genetics*, 169(4), 1807–14. doi:10.1534/genetics.104.035352
- Kusumoto, M., Suzuki, R., Nishiya, Y., Okitsu, T., & Oka, M. (2004). Host-Dependent Activation of IS 1203 v Excision in Shiga Toxin-Producing Escherichia coli, 97(6), 406–411.
- Laeremans, T., Coolsaet, N., Verreth, C., Snoeck, C., Hellings, N., Vanderleyden, J., & Martinez-Romero, E. (1997). Functional redundancy of genes for sulphate activation enzymes in Rhizobium sp. BR816. *Microbiology*, 143(12), 3933–3942. doi:10.1099/00221287-143-12-3933
- Laing, C., Buchanan, C., Taboada, E. N., Zhang, Y., Kropinski, A., Villegas, A., ... Gannon, V. P. J. (2010). Pan-genome sequence analysis using Panseq: an

online tool for the rapid analysis of core and accessory genomic regions. *BMC Bioinformatics*, 11, 461. doi:10.1186/1471-2105-11-461

- Lam SD., Das S., Sillitoe I., Orengo C.  
An overview of comparative modelling and resources dedicated to large-scale modelling of genome sequences. *Acta Crystallogr D Struct Biol.* 2017 Aug 1;73(Pt 8):628-640. doi: 10.1107/S2059798317008920. Epub 2017 Jul 28.
- Laskowski, R. a., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*, 26(2), 283–291. doi:10.1107/S0021889892009944
- Little, J. W., Mount, D. W., & Yanisch-Perron, C. R. (1981). Purified *lexA* protein is a repressor of the *recA* and *lexA* genes. *Proceedings of the National Academy of Sciences of the United States of America*, 78(7), 4199–203.
- Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., ... Law, M. (2012). Comparison of next-generation sequencing systems. *Journal of Biomedicine & Biotechnology*, 2012, 251364. doi:10.1155/2012/251364
- Lloyd, R. G. (1991). Conjugational recombination in resolvase-deficient *ruvC* mutants of *Escherichia coli* K-12 depends on *recG*. *Journal of Bacteriology*, 173(17), 5414–8.
- Lockyer, J., Cook, R. G., Milstien, S., Kaufman, S., Woo, S. L., & Ledley, F. D. (1987). Structure and expression of human dihydropteridine reductase. *Proceedings of the National Academy of Sciences of the United States of America*, 84(10), 3329–33.
- Loughlin, M. F., Barnard, F. M., Jenkins, D., Sharples, J., Jenks, P. J., & Sharples, G. J. (2003). *Helicobacter pylori* Mutants Defective in RuvC Holliday Junction Resolvase Display Reduced Macrophage Survival and Spontaneous Clearance from the Murine Gastric Mucosa *Helicobacter pylori* Mutants Defective in RuvC Holliday Junction Resolvase Display Reduce. doi:10.1128/IAI.71.4.2022
- Luo, Y., Pfuetzner, R. A., Mosimann, S., Paetzel, M., Frey, E. A., Cherney, M., ... Strynadka, N. C. J. (2001). Crystal Structure of LexA : A Conformational Switch for Regulation of Self-Cleavage, 106, 585–594.
- Machida, C., & Machida, Y. (1989). Regulation of IS1 transposition by the *insA* gene product. *Journal of Molecular Biology*, 208(4), 567–74.
- Mahdi, a a, Sharples, G. J., Mandal, T. N., & Lloyd, R. G. (1996). Holliday junction resolvases encoded by homologous *rusA* genes in *Escherichia coli* K-12 and phage 82. *Journal of Molecular Biology*, 257(3), 561–73. doi:10.1006/jmbi.1996.0185

- Mahillon, J., & Chandler, M. (1998). Insertion Sequences Insertion Sequences, *62*(3).
- Mandal, T. N., Mahdi, A. A., Sharples, G. J., & Lloyd, R. G. (1993). Resolution of Holliday intermediates in recombination and DNA repair: indirect Resolution of Holliday Intermediates in Recombination and DNA Repair: Indirect Suppression of *ruvA*, *ruvB*, and *ruvC* Mutations.
- Maskell, J. P., Sefton, A. M., Hall, L. M. C., Ott, M., Roeder, D., Stuber, D., ... Wipf, B. (2001). Multiple Mutations Modulate the Function of Dihydrofolate Reductase in Trimethoprim-Resistant *Streptococcus pneumoniae*, *45*(4), 1104–1108. doi:10.1128/AAC.45.4.1104
- Maurelli, A. T. (2007). Black holes, antivirulence genes, and gene inactivation in the evolution of bacterial pathogens. *FEMS Microbiology Letters*, *267*(1), 1–8. doi:10.1111/j.1574-6968.2006.00526.x
- Mccormick, M. L., Buettner, G. R., Britigan, B. E., & Cormick, M. L. M. C. (1998). Endogenous Superoxide Dismutase Levels Regulate Iron-Dependent Hydroxyl Radical Formation in *Escherichia coli* Exposed to Hydrogen Peroxide Endogenous Superoxide Dismutase Levels Regulate Iron-Dependent Hydroxyl Radical Formation in *Escherichia coli* Expose.
- McCrea, K. W., Wang, M. L., Xie, J., Sandstedt, S. a, Davis, G. S., Lee, J. H., ... Gilsdorf, J. R. (2010). Prevalence of the *sodC* gene in nontypeable *Haemophilus influenzae* and *Haemophilus haemolyticus* by microarray-based hybridization. *Journal of Clinical Microbiology*, *48*(3), 714–9. doi:10.1128/JCM.01416-09
- Mira, A., Martín-Cuadrado, A. B., D'Auria, G., & Rodríguez-Valera, F. (2010). The bacterial pan-genome: a new paradigm in microbiology. *International Microbiology: The Official Journal of the Spanish Society for Microbiology*, *13*(2), 45–57. doi:10.2436/20.1501.01.110
- Muniesa, M., Hammerl, J. a, Hertwig, S., Appel, B., & Brüßow, H. (2012). Shiga toxin-producing *Escherichia coli* O104:H4: a new challenge for microbiology. *Applied and Environmental Microbiology*, *78*(12), 4065–73. doi:10.1128/AEM.00217-12
- Nakao, R., Ramstedt, M., Wai, S. N., & Uhlin, B. E. (2012). Enhanced biofilm formation by *Escherichia coli* LPS mutants defective in Hep biosynthesis. *PLoS One*, *7*(12), e51241. doi:10.1371/journal.pone.0051241
- Nare, B. (1997). The Roles of Pteridine Reductase 1 and Dihydrofolate Reductase-Thymidylate Synthase in Pteridine Metabolism in the Protozoan Parasite *Leishmania major*. *Journal of Biological Chemistry*, *272*(21), 13883–13891. doi:10.1074/jbc.272.21.13883
- Niba, E. T. E., Naka, Y., Nagase, M., Mori, H., & Kitakawa, M. (2007). A genome-wide approach to identify the genes involved in biofilm formation in *E. coli*. *DNA Research: An International Journal for Rapid Publication of Reports on Genes and Genomes*, *14*(6), 237–46. doi:10.1093/dnares/dsm024

- Nichols, R. J., Sen, S., Choo, Y. J., Beltrao, P., Zietek, M., Chaba, R., ... Gross, C. a. (2011). Phenotypic landscape of a bacterial cell. *Cell*, *144*(1), 143–56. doi:10.1016/j.cell.2010.11.052
- Omelchenko, M. V, Galperin, M. Y., Wolf, Y. I., & Koonin, E. V. (2010). Non-homologous isofunctional enzymes: a systematic analysis of alternative solutions in enzyme evolution. *Biology Direct*, *5*, 31. doi:10.1186/1745-6150-5-31
- Ooka, T., Ogura, Y., Asadulghani, M., Ohnishi, M., Nakayama, K., Terajima, J., ... Hayashi, T. (2009). Inference of the impact of insertion sequence (IS) elements on bacterial genome diversification through analysis of small-size structural polymorphisms in *Escherichia coli* O157 genomes. *Genome Research*, *19*(10), 1809–16. doi:10.1101/gr.089615.108
- Otsuji, N., Iyehara, H., & Hideshima, Y. (1974). Isolation and characterization of an *Escherichia coli* *ruv* mutant which forms nonseptate filaments after low doses of ultraviolet light irradiation. *Journal of Bacteriology*, *117*(2), 337–44.
- Otto, T. D., Guimarães, A. C. R., Degraeve, W. M., & de Miranda, A. B. (2008). AnEnPi: identification and annotation of analogous enzymes. *BMC Bioinformatics*, *9*, 544. doi:10.1186/1471-2105-9-544
- Pacello, F., Ceci, P., Ammendola, S., Pasquali, P., Chiancone, E., & Battistoni, A. (2008). Periplasmic Cu,Zn superoxide dismutase and cytoplasmic Dps concur in protecting *Salmonella enterica* serovar Typhimurium from extracellular reactive oxygen species. *Biochimica et Biophysica Acta*, *1780*(2), 226–32. doi:10.1016/j.bbagen.2007.12.001
- Pallen, M. J., & Wren, B. W. (2007). Bacterial pathogenomics. *Nature*, *449*(7164), 835–42. doi:10.1038/nature06248
- Pappachan, a, Savithri, H. S., & Murthy, M. R. N. (2008a). Structural and functional studies on a mesophilic stationary phase survival protein (Sur E) from *Salmonella typhimurium*. *The FEBS Journal*, *275*(23), 5855–64. doi:10.1111/j.1742-4658.2008.06715.x
- Pappachan, a, Savithri, H. S., & Murthy, M. R. N. (2008b). Structural and functional studies on a mesophilic stationary phase survival protein (Sur E) from *Salmonella typhimurium*. *The FEBS Journal*, *275*(23), 5855–64. doi:10.1111/j.1742-4658.2008.06715.x
- Park, P. J. (2009). ChIP-seq: advantages and challenges of a maturing technology. *Nature Reviews. Genetics*, *10*(10), 669–80. doi:10.1038/nrg2641
- Patel, C. N., Wortham, B. W., Lines, J. L., Fetherston, J. D., Perry, R. D., Marcos, A., & Oliveira, M. A. (2006). Polyamines Are Essential for the Formation of Plague Biofilm Polyamines Are Essential for the Formation of Plague Biofilm. doi:10.1128/JB.188.7.2355

- Pearson, C. L., Loshon, C. A., Pedersen, L. B., Setlow, B., & Setlow, P. (2000). Analysis of the Function of a Putative Phosphoglycerate Mutase from *Bacillus subtilis*. *Journal of Bacteriology*, *182*(14), 4121–4123. doi:10.1128/JB.182.14.4121-4123.2000.Updated
- Phosphohydrolyases, A., & Selengut, J. D. (2001). MDP-1 Is a New and Distinct Member of the Haloacid Dehalogenase Family of, *Journal of Bacteriology*, *183*(1), 12704–12711.
- Piddington, D. L., Fang, F. C., Laessig, T., Cooper, a M., Orme, I. M., & Buchmeier, N. a. (2001). Cu,Zn superoxide dismutase of *Mycobacterium tuberculosis* contributes to survival in activated macrophages that are generating an oxidative burst. *Infection and Immunity*, *69*(8), 4980–7. doi:10.1128/IAI.69.8.4980-4987.2001
- Pribat, A., Blaby, I. K., Lara-Núñez, A., Gregory, J. F., de Crécy-Lagard, V., & Hanson, A. D. (2010). FolX and FolM are essential for tetrahydromonapterin synthesis in *Escherichia coli* and *Pseudomonas aeruginosa*. *Journal of Bacteriology*, *192*(2), 475–82. doi:10.1128/JB.01198-09
- Proudfoot, M., Kuznetsova, E., Brown, G., Rao, N. N., Kitagawa, M., Mori, H., ... Yakunin, A. F. (2004). General enzymatic screens identify three new nucleotidases in *Escherichia coli*. Biochemical characterization of SurE, YfbR, and YjjG. *The Journal of Biological Chemistry*, *279*(52), 54687–94. doi:10.1074/jbc.M411023200
- Quackenbush, J. (2002). Microarray data normalization and transformation. *Nature Genetics*, *32 Suppl*(december), 496–501. doi:10.1038/ng1032
- Rafferty, J. B., Bolt, E. L., Muranova, T. a, Sedelnikova, S. E., Leonard, P., Pasquo, A., ... Lloyd, R. G. (2003). The Structure of *Escherichia coli* RusA Endonuclease Reveals a New Holliday Junction DNA Binding Fold. *Structure*, *11*(12), 1557–1567. doi:10.1016/j.str.2003.11.004
- Raghava, G. P. S., & Han, J. H. (2005). Correlation and prediction of gene expression level from amino acid and dipeptide composition of its protein. *BMC Bioinformatics*, *6*, 59. doi:10.1186/1471-2105-6-59
- Rasko, D. a, Rosovitz, M. J., Myers, G. S. a, Mongodin, E. F., Fricke, W. F., Gajer, P., ... Ravel, J. (2008a). The pangenome structure of *Escherichia coli*: comparative genomic analysis of *E. coli* commensal and pathogenic isolates. *Journal of Bacteriology*, *190*(20), 6881–93. doi:10.1128/JB.00619-08
- Rasko, D. a, Rosovitz, M. J., Myers, G. S. a, Mongodin, E. F., Fricke, W. F., Gajer, P., ... Ravel, J. (2008b). The pangenome structure of *Escherichia coli*: comparative genomic analysis of *E. coli* commensal and pathogenic isolates. *Journal of Bacteriology*, *190*(20), 6881–93. doi:10.1128/JB.00619-08
- Robu, M. E., Inman, R. B., & Cox, M. M. (n.d.). RecA protein promotes the regression of stalled replication forks in vitro.

- Rump, L. V, Fischer, M., & González-Escalona, N. (2011). Different IS629 transposition frequencies exhibited by *Escherichia coli* O157:H7 strains in the stepwise evolutionary model. *Applied and Environmental Microbiology*, 77(14), 5030–3. doi:10.1128/AEM.00249-11
- Saenger, A. (1999). X-ray structure of the *Escherichia coli* periplasmic 5' - nucleotidase containing, 6(5).
- Saier, M. H., & Zhang, Z. (2014). Transposon-mediated directed mutation controlled by DNA binding proteins in *Escherichia coli*. *Frontiers in Microbiology*, 5(August), 390. doi:10.3389/fmicb.2014.00390
- Sanjay, M. K., Shrideshikan, S. M., Usha, M. S., Philipraj, a, Gaddad, S. M., & Shivannavar, C. T. (2010). Detection, amplification & sequence homology of *sodC* in clinical isolates of *Salmonella* sp. *The Indian Journal of Medical Research*, 131(April), 565–70.
- Sansone, A., Watson, P. R., Wallis, T. S., Langford, P. R., & Kroll, J. S. (2002). The role of two periplasmic copper- and zinc-cofactored superoxide dismutases in the virulence of *Salmonella choleraesuis*, 719–726.
- Santos, C. A., Saraiva, A. M., Toledo, M. A. S., Beloti, L. L., Crucello, A., Favaro, M. T. P., ... Souza, A. P. (2013). Microbial Pathogenesis Initial biochemical and functional characterization of a 5' 0 -nucleotidase from *Xylella fastidiosa* related to the human cytosolic 5' 0 -nucleotidase I. *Microbial Pathogenesis*, 59-60, 1–6. doi:10.1016/j.micpath.2013.02.007
- Schmidt, H., & Hensel, M. (2004). Pathogenicity Islands in Bacterial Pathogenesis Pathogenicity Islands in Bacterial Pathogenesis, 17(1). doi:10.1128/CMR.17.1.14
- Schnell, J. R., Dyson, H. J., & Wright, P. E. (2004). Structure, dynamics, and catalytic function of dihydrofolate reductase. *Annual Review of Biophysics and Biomolecular Structure*, 33, 119–40. doi:10.1146/annurev.biophys.33.110502.133613
- Schnetz, K., & Rak, B. (1992). IS5: a mobile enhancer of transcription in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America*, 89(4), 1244–8.
- Sharples, G. J., Benson, F. E., Illing, G. T., & Lloyd, R. G. (1990). Molecular and functional analysis of the *ruv* region of *Escherichia coli* K-12 reveals three genes involved in DNA repair and recombination. *Molecular & General Genetics : MGG*, 221(2), 219–26.
- Shinagawa, H., & Iwasaki, H. (1996). Processing the Holliday junction, 0004(March), 107–111.

- Smith, D. R., & Chapman, M. R. (2010). Economical Evolution : Microbes Economical Evolution : Microbes Reduce the Synthetic Cost of. doi:10.1128/mBio.00131-10.Updated
- Soo, V. W. C., Hanson-Manful, P., & Patrick, W. M. (2011). Artificial gene amplification reveals an abundance of promiscuous resistance determinants in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(4), 1484–9. doi:10.1073/pnas.1012108108
- Sperandeo, P., Pozzi, C., Dehò, G., & Polissi, A. (2006). Non-essential KDO biosynthesis and new essential cell envelope biogenesis genes in the *Escherichia coli* yrbG-yhbG locus. *Research in Microbiology*, *157*(6), 547–58. doi:10.1016/j.resmic.2005.11.014
- Smith HO, Danner DB, Deich RA. (1981); Genetic transformation. Review. *Annu Rev Biochem.* 50:41-68.
- Spoering, A. L., Vulic, M., & Lewis, K. (2006). GlpD and PlsB participate in persister cell formation in *Escherichia coli*. *Journal of Bacteriology*, *188*(14), 5136–44. doi:10.1128/JB.00369-06
- Stallings, W. C., Powers, T. B., Pattridge, K. A., Fee, J. A., & Ludwig, M. L. (1983). *Iron superoxide dismutase from Escherichia coli at 3.1-Å resolution: A structure unlike that of copper / zinc protein at both monomer and dimer levels.*
- Steiert, P. S., Stauffer, L. T., & Stauffer, G. V. (1990). The lpd gene product functions as the L protein in the *Escherichia coli* glycine cleavage enzyme The lpd Gene Product Functions as the L Protein in the *Escherichia coli* Glycine Cleavage Enzyme System, *172*(10), 10–13.
- Sub classification and targeted characterization of prophage-encoded two-component cell lysis cassette. (2007), *32*(August).
- T., H., a., N., M., V.-G., & T., H. (2001). Site-directed mutagenesis and phylogenetic comparisons of the *Escherichia coli* Tus protein: DNA-protein interactions alone can not account for Tus activity. *Molecular Genetics and Genomics*, *265*(6), 941–953. doi:10.1007/s004380100501
- Tan, K. A. I. S. O. O., Wee, B. Y. U., & Song, K. P. (2001). Evidence for holin function of tcdE gene in the pathogenicity of *Clostridium dif* ® cile, *50*, 613–619.
- Templin, M. F., Stoll, D., Schrenk, M., Traub, P. C., Vöhringer, C. F., & Joos, T. O. (2002). Protein microarray technology, *20*(4), 160–166.
- Tettelin, H., Riley, D., Cattuto, C., & Medini, D. (2008). Comparative genomics: the bacterial pan-genome. *Current Opinion in Microbiology*, *11*(5), 472–477. doi:10.1016/j.mib.2008.09.006

- Thammavongsa, V., Kern, J. W., & Missiakas, D. M. (2009). Staphylococcus aureus synthesizes adenosine to escape host immune responses, *206*(11), 2417–2427. doi:10.1084/jem.20090097
- Toba, F. a, Thompson, M. G., Campbell, B. R., Junker, L. M., Rueggeberg, K.-G., & Hay, A. G. (2011a). Role of DLP12 lysis genes in Escherichia coli biofilm formation. *Microbiology (Reading, England)*, *157*(Pt 6), 1640–50. doi:10.1099/mic.0.045161-0
- Toba, F. a, Thompson, M. G., Campbell, B. R., Junker, L. M., Rueggeberg, K.-G., & Hay, A. G. (2011b). Role of DLP12 lysis genes in Escherichia coli biofilm formation. *Microbiology (Reading, England)*, *157*(Pt 6), 1640–50. doi:10.1099/mic.0.045161-0
- Toprak, E., Veres, A., Michel, J., Chait, R., Hartl, D. L., & Kishony, R. (2011). Evolutionary paths to antibiotic resistance under dynamically sustained drug selection. *Nature Genetics*, *44*(1), 101–105. doi:10.1038/ng.1034
- Touchon, M., Hoede, C., Tenaillon, O., Barbe, V., Baeriswyl, S., Bidet, P., ... Denamur, E. (2009). Organised genome dynamics in the Escherichia coli species results in highly diverse adaptive paths. *PLoS Genetics*, *5*(1), e1000344. doi:10.1371/journal.pgen.1000344
- Tsai, M. M., Wong, R. Y., Hoang, a T., & Deonier, R. C. (1987). Transposition of Tn1000: in vivo properties. *Journal of Bacteriology*, *169*(12), 5556–62.
- Uzzau, S., Bossi, L., & Figueroa-Bossi, N. (2002). Differential accumulation of Salmonella[Cu, Zn] superoxide dismutases SodCI and SodCII in intracellular bacteria: correlation with their relative contribution to pathogenicity. *Molecular Microbiology*, *46*(1), 147–156. doi:10.1046/j.1365-2958.2002.03145.x
- Vanaporn, M., Wand, M., Michell, S. L., Sarkar-Tyson, M., Ireland, P., Goldman, S., ... Titball, R. W. (2011). Superoxide dismutase C is required for intracellular survival and virulence of Burkholderia pseudomallei. *Microbiology (Reading, England)*, *157*(Pt 8), 2392–400. doi:10.1099/mic.0.050823-0
- Varughese, K. I., Skinner, M. M., Whiteley, J. M., Matthews, D. a, & Xuong, N. H. (1992). Crystal structure of rat liver dihydropteridine reductase. *Proceedings of the National Academy of Sciences of the United States of America*, *89*(13), 6080–4.
- Verwoert, I. I., Verbree, E. C., Linden, K. H. Van Der, Nijkamp, H. J., & Stuitje, R. (1992). malonyl coenzyme A-acyl carrier protein Cloning , Nucleotide Sequence , and Expression of the Escherichia coli fabD Gene , Encoding Malonyl Coenzyme A-Acyl Carrier Protein Transacylase.
- View, S. E. A. (1996). SEA VIEW and PHYLO\_ WIN: two graphic tools for sequence alignment and molecular phylogeny, *12*(6), 543–548.

- Von Mering, C., Jensen, L. J., Snel, B., Hooper, S. D., Krupp, M., Foglierini, M., ... Bork, P. (2005). *STRING* : known and predicted protein-protein associations, integrated and transferred across organisms. *Nucleic Acids Research*, 33(Database issue), D433–7. doi:10.1093/nar/gki005
- Wang, I. N., Smith, D. L., & Young, R. (2000). Holins: the protein clocks of bacteriophage infections. *Annual Review of Microbiology*, 54, 799–825. doi:10.1146/annurev.micro.54.1.799
- Wang, J., Leblanc, E., Chang, C. F., Papadopoulou, B., Bray, T., Whiteley, J. M., ... Ouellette, M. (1997). Pterin and folate reduction by the *Leishmania tarentolae* H locus short-chain dehydrogenase/reductase PTR1. *Archives of Biochemistry and Biophysics*, 342(2), 197–202. doi:10.1006/abbi.1997.0126
- Wang, X., Kim, Y., Ma, Q., Hong, S. H., Pokusaeva, K., Sturino, J. M., & Wood, T. K. (2010). Cryptic prophages help bacteria cope with adverse environments. *Nature Communications*, 1(9), 147. doi:10.1038/ncomms1146
- Weissborn, A. C., & Kennedy, P. (1984). Biosynthesis of Membrane-derived Oligosaccharides, 12644–12651.
- West, S. C. (1997). Processing of recombination intermediates by the RuvABC proteins. *Annual Review of Genetics*, 31, 213–44. doi:10.1146/annurev.genet.31.1.213
- Wilks, K. E., Dunn, K. L. R., Farrant, J. L., Karen, M., Gorringer, A. R., Langford, P. R., ... Kroll, J. S. (1998). Periplasmic Superoxide Dismutase in Meningococcal Pathogenicity Periplasmic Superoxide Dismutase in Meningococcal Pathogenicity.
- Wu, W. H., Morris, D. R., & Morris, R. (1973). ARTICLE: Biosynthetic Arginine Decarboxylase from *Escherichia coli*: PURIFICATION AND PROPERTIES Decarboxylase from *Escherichia coli*.
- Yamada, K., Ariyoshi, M., & Morikawa, K. (2004). Three-dimensional structural views of branch migration and resolution in DNA homologous recombination. *Current Opinion in Structural Biology*, 14(2), 130–7. doi:10.1016/j.sbi.2004.03.005
- Yoshimura, M., Oshima, T., & Ogasawara, N. (2007). Involvement of the YneS/YgiH and PlsX proteins in phospholipid biosynthesis in both *Bacillus subtilis* and *Escherichia coli*. *BMC Microbiology*, 7(Figure 1), 69. doi:10.1186/1471-2180-7-69
- Young, K., & Silver, L. L. (1991). strains of *Escherichia coli* . Leakage of Periplasmic Enzymes from envA1 Strains of *Escherichia coli*, 173(12).
- Young, R. Y. (1992). Bacteriophage Lysis : Mechanism and Regulation, 56(3).
- Zagursky, R. J., Ooi, P., Jones, K. F., Fiske, M. J., Smith, R. P., Green, B. A., ... York, N. (2000). Identification of a *Haemophilus influenzae* 5 J -Nucleotidase

Protein : Cloning of the nucA Gene and Immunogenicity and Characterization of the NucA Protein, 68(5), 2525–2534.

- Zhang, G., Morais, M. C., Dai, J., Zhang, W., Dunaway-Mariano, D., & Allen, K. N. (2004). Investigation of metal ion binding in phosphonoacetaldehyde hydrolase identifies sequence markers for metal-activated enzymes of the HAD enzyme superfamily. *Biochemistry*, 43(17), 4990–7. doi:10.1021/bi036309n
- Zhang, J. W., Butland, G., Greenblatt, J. F., Emili, A., & Zamble, D. B. (2005). A role for SlyD in the Escherichia coli hydrogenase biosynthetic pathway. *The Journal of Biological Chemistry*, 280(6), 4360–6. doi:10.1074/jbc.M411799200
- Zimmerman, M. D., Proudfoot, M., Yakunin, A., & Minor, W. (2008). Structural insight into the mechanism of substrate specificity and catalytic activity of an HD-domain phosphohydrolase: the 5'-deoxyribonucleotidase YfbR from Escherichia coli. *Journal of Molecular Biology*, 378(1), 215–26. doi:10.1016/j.jmb.2008.02.036