UNIVERSIDADE FEDERAL DE GOIÁS
Instituto de Ciências Biológicas
Programa de Pós-Graduação em Genética
e Biologia Molecular

KARLA CHRISTINA SOUSA SILVA

ANÁLISES PROTEÔMICAS DE CEPAS DE *Staphylococcus saprophyticus* ELUCIDAM DIFERENTES ESTRATÉGIAS PARA VIRULÊNCIA E INFECÇÃO

Goiânia
2019

UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS

TERMO DE CIÊNCIA E DE AUTORIZAÇÃO (TECA) PARA DISPONIBILIZAR VERSÕES ELETRÔNICAS DE TESES

E DISSERTAÇÕES NA BIBLIOTECA DIGITAL DA UFG

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás (UFG) a disponibilizar, gratuitamente, por meio da Biblioteca Digital de Teses e Dissertações (BDTD/UFG), regulamentada pela Resolução CEPEC nº 832/2007, sem ressarcimento dos direitos autorais, de acordo com a Lei 9.610/98, o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou download, a título de divulgação da produção científica brasileira, a partir desta data.

O conteúdo das Teses e Dissertações disponibilizado na BDTD/UFG é de responsabilidade exclusiva do autor. Ao encaminhar o produto final, o autor(a) e o(a) orientador(a) firmam o compromisso de que o trabalho não contém nenhuma violação de quaisquer direitos autorais ou outro direito de terceiros.

1. Identificação do material bibliográfico

[ ] Dissertação [x] Tese

2. Nome completo do autor

Karla Christina Sousa Silva

3. Título do trabalho

Análises proteômicas de cepas de Staphylococcus saprophyticus elucidam diferentes estratégias para virulência e infecção

4. Informações de acesso ao documento (este campo deve ser preenchido pelo orientador)

Concorda com a liberação total do documento [x] SIM [ ] NÃO[1]

[1] Neste caso o documento será embargado por até um ano a partir da data de defesa. Após esse período, a possível disponibilização ocorrerá apenas mediante:
a) consulta ao(à) autor(a) e ao(à) orientador(a);
b) novo Termo de Ciência e de Autorização (TECA) assinado e inserido no arquivo da tese ou dissertação.
O documento não será disponibilizado durante o período de embargo.
Casos de embargo:
- Solicitação de registro de patente;
- Submissão de artigo em revista científica;
- Publicação como capítulo de livro;
- Publicação da dissertação/tese em livro.

Obs. Este termo deverá ser assinado no SEI pelo orientador e pelo autor.

Documento assinado eletronicamente por **Juliana Alves Parente Rocha**, **Professora do Magistério Superior**, em 14/06/2022, às 11:51, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Karla Christina Sousa Silva**, **Usuário Externo**, em 20/06/2022, às 13:02, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **2978620** e o código CRC **5CEC6B72**.

Referência: Processo nº 23070.025127/2021-46 SEI nº 2978620

KARLA CHRISTINA SOUSA SILVA

ANÁLISES PROTEÔMICAS DE CEPAS DE *Staphylococcus saprophyticus* ELUCIDAM DIFERENTES ESTRATÉGIAS PARA VIRULÊNCIA E INFECÇÃO

Tese apresentada ao Programa de Pós-Graduação em Genética e Biologia Molecular do Instituto de Ciências Biológicas da Universidade Federal de Goiás (UFG), como requisito para obtenção do título de Doutora em Genética e Biologia Molecular.
Área de concentração: Genética e Biologia Molecular.
Linha de pesquisa: Proteômica da interação patógeno-hospedeiro.
Orientadora: Dra. Juliana Alves Parente-Rocha.

Goiânia

2019

Ficha de identificação da obra elaborada pelo autor, através do
Programa de Geração Automática do Sistema de Bibliotecas da UFG.

Silva, Karla Christina Sousa
ANÁLISES PROTEÔMICAS DE CEPAS DE Staphylococcus saprophyticus ELUCIDAM DIFERENTES ESTRATÉGIAS PARA VIRULÊNCIA E INFECÇÃO [manuscrito] / Karla Christina Sousa Silva - 2019.
XCVII, 97 f.: il.

Orientador: Prof. Dr. Juliana Alves Parente-Rocha.
Tese (Doutorado) - Universidade Federal de Goiás, Instituto de Ciências Biológicas (ICB), Programa de Pós-Graduação em Genética e Biologia Molecular, Goiânia, 2019.
Bibliografia. Anexos.

1. Proteômica. 2. Staphylococcus saprophyticus. 3. Virulência. 4. Infecção. 5. Interação patógeno-hospedeiro. I. Parente-Rocha, Juliana Alves, orient. II. Título.

CDU 577.2

SERVIÇO PÚBLICO FEDERAL
UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E BIOLOGIA MOLECULAR

ATA DA SESSÃO PÚBLICA DE DEFESA DE TESE DE Nº 018

Aos treze dias do mês de maio do ano de dois mil e dezenove (13/05/2019), às 09hs00min, no Auditório do ICB V, reuniram-se os componentes da banca examinadora: Profa. Dra. Juliana Alves Parente-Rocha, orientadora ; Prof. Dr. Clayton Luiz Borges; Prof. Dr. André Kipnis; Profa. Dra. Márcia Giambiagi de Marval e Prof. Dr. Wagner Fontes, para, em sessão pública presidida pela primeira examinadora citada, procederem à avaliação da defesa de tese intitulada: **"Análises proteômicas de cepas de *Staphylococcus saprophyticus* elucidam diferentes estratégias para virulência e infecção"**, em nível de doutorado, área de concentração em **Genética e Biologia Molecular**, de autoria de **Karla Christina Sousa Silva**, discente do Programa de Pós-Graduação em Genética e Biologia Molecular da Universidade Federal de Goiás. A sessão foi aberta pela presidente da banca, Profa. Dra. Juliana Alves Parente-Rocha, que fez a apresentação formal dos membros da banca. A palavra, a seguir, foi concedida à autora da tese que, em cerca de 40 minutos, procedeu à apresentação de seu trabalho. Terminada a apresentação, cada membro da banca arguiu a examinada, tendo-se adotado o sistema de diálogo sequencial. Terminada a fase de arguição, procedeu-se a avaliação da tese. Tendo-se em vista o que consta na Resolução nº 1294 de 06 de Junho de 2014 do Conselho de Ensino, Pesquisa, Extensão e Cultura (CEPEC), que regulamenta o Programa de Pós-Graduação em Genética e Biologia Molecular, a tese foi APROVADA, considerando-se integralmente cumprido este requisito para fins de obtenção do título de Doutor (a) em Genética e Biologia Molecular pela Universidade Federal de Goiás. A conclusão do curso dar-se-á quando da entrega da versão definitiva da tese na secretaria do programa, com as devidas correções sugeridas pela banca examinadora, no prazo de trinta dias a contar da data da defesa. Cumpridas

SERVIÇO PÚBLICO FEDERAL
UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E BIOLOGIA MOLECULAR

as formalidades de pauta, às 12 horas e 50 minutos, encerrou-se a sessão de defesa e, para constar, a presidente da Banca lavrou a presente ata que, após lida e aprovada, será assinada pelos membros da banca examinadora em três vias de igual teor.

Juliana A. P. Rocha

Profa. Dra. Juliana Alves Parente-Rocha
Presidente da Banca
ICB
UFG/GO

Prof. Dr. Clayton Luiz Borges
ICB
UFG/GO

Prof. Dr. André Kipnis
UFG/GO

Profa. Dra. Márcia Giambiagi de Marval
UFRJ/RJ

Prof. Dr. Wagner Fontes
UNB/DF

3CARTÓRIO
DE REGISTRO CIVIL E TABELIONATO DE NOTAS
62 3229-3097
AUTENTICAÇÃO
Confere com o original. 03 de dezembro de 2019.
Danilo Junio Laureano dos Santos
Escrevente
06961912010194409490854
https://see.tjgo.jus.br/buscas

SERVIÇO PÚBLICO FEDERAL
UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E BIOLOGIA MOLECULAR

SESSÃO PÚBLICA DE DEFESA DE TESE Nº 018

TÍTULO DO TRABALHO: "**Análises proteômicas de cepas de *Staphylococcus saprophyticus* elucidam diferentes estratégias para virulência e infecção**".

AUTOR (A): Karla Christina Sousa Silva

TESE Nº 018 DO PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E BIOLOGIA MOLECULAR, DEFENDIDA E APROVADA EM SESSÃO PÚBLICA, NA DATA DE 13/05/2019, ÀS 09h00min, no Auditório do ICB V, CUJA BANCA EXAMINADORA ESTEVE CONSTITUÍDA DOS SEGUINTES PROFESSORES:

Profa. Dra. Juliana Alves Parente-Rocha ICB UFG/GO	
Prof. Dr. Clayton Luiz Borges ICB UFG/GO	Prof. Dr. André Kipnis UFG/GO
Profa. Dra. Márcia Giambiagi de Marval UFRJ/RJ	Prof. Dr. Wagner Fontes UNB/DF

HOUVE MUDANÇA NO TÍTULO? SIM ( ) NÃO ( )

CASO HAJA MUDANÇA, ESCREVER ABAIXO O NOVO TÍTULO DA DISSERTAÇÃO:

.....

.....

.....

OBS: O título na dissertação impressa deve corresponder exatamente ao que se encontra na ata.

GOIÂNIA (GO) 13/05/2019

Dedicatória

Dedico a aquele que era para ter sido o primeiro: Alex Jesus Carvalho *in memorian.*

"*That's what I do, I drink and I know things."*

Tyrion Lannister, *Game of Thrones*, Temporada 6 – Episódio 2

Sumário

RESUMO

Staphylococcus saprophyticus é uma bactéria Gram-positiva, coagulase negativa e é um dos agentes etiológicos mais comuns de infecções do trato urinário (ITU) em mulheres jovens sexualmente ativas. A infecção pode ser adquirida principalmente via relação sexual. Pouco se sabe sobre suas características moleculares no modelo de *S. saprophyticus*. No entanto, a plasticidade no perfil proteômico é detectada no gênero *Staphylococcus* e pode estar relacionada com a capacidade de infecção. Neste estudo, objetivou-se analisar as diferenças metabólicas entre três diferentes cepas patogênicasde *S. saprophyticus* (ATCC 15305, 7108 e 9325) utilizando abordagens proteômicas. Os dados proteômicos revelaram variações entre as cepas que podem levar a diferentes respostas no contexto de uma infecção. Existem diferenças no metabolismo de carboidratos, no metabolismo de aminoácidos e na defesa de patógenos contra as defesas do hospedeiro. A tiorredoxina foi uma das proteínas diferencialmente expressas entre as cepas; os níveis de tiol mostraram níveis semelhantes de tiol reduzido para as cepas ATCC15305 e 9325; em contraste, a cepa 7108 mostrou níveis consideravelmente baixos de níveis tiol reduzidos quando comparados com as outras cepas analisadas. A urease é um dos principais fatores de virulência do *S. saprophyticus*. As concentrações de urease citoplasmática variaram consideravelmente. A linhagem 9325 apresentou a menor atividade da urease citoplasmática, porém quando aplicada em ágar à base de uréia, apresenta alta atividade de urease, o que significa que provavelmente a maior parte da urease sintetizada é exportada para o meio extracelular. As cepas também foram analisadas. Outra característica importante está relacionada ao operon do metabolismo das purinas, que é altamente regulado na ATCC15305 e exerce uma influência sobre a produção de biofilme nesta linhagem. Estes resultados mostram que, apesar de pertencerem à mesma espécie, diferentes linhagens de *S. saprophyticus* apresentam comportamento diverso em resposta a diferentes contextos que o microrganismo pode enfrentar no hospedeiro.

Palavras-chave: Proteômica, *Staphylococcus saprophyticus*, virulência, infecção.

ABSTRACT

Staphylococcus saprophyticus is a gram positive, coagulase negative bacteria and is one of the most common etiological agent of urinary tract infections (UTI) among sexually active young females. The infection may be led mainly by sexual intercourse. Little is known about the molecular features of *S. saprophyticus* infection model. However, plasticity in the proteomic profile is detected in the genus *Staphylococcus* and can be related with infection ability. In this study, we aimed to analyze the metabolic differences among three different pathogenic strains of *S. saprophyticus* (ATCC 15305, 7108 and 9325) using proteomics approaches. The proteomic data revealed variations among the strains that may lead to different responses in the context of an infection. There are differences in the carbohydrate metabolism, amino acid metabolism and pathogen defense against host defenses. Thioredoxin was one of the differentially expressed proteins among strains, thiol level assays showed similar levels of reduced thiol for ATCC15305 and 9325 strains, in contrast 7108 strain showed considerably low levels of reduced thiol levels when compared to the other analyzed strains. Urease is one of the main virulence factors of *S. saprophyticus*. The cytoplasmatic urease concentrations varied considerably. The 9325 strain had the lowest cytoplasmatic urease activity, however when plated on urea-based agar it showed a high urease activity, which means that probably most of the urease synthesized is exported to the extracellular milieu. Another important feature is related to the purine metabolism operon, which is highly regulated on ATCC15305 and exerts an influence upon the biofilm production in this strain. These results show that, although belonging to the same species, different *S. saprophyticus* strains present diverse behavior in response to different contexts that the microorganism may face in the host.

Key-words: Proteomics, *Staphylococcus saprophyticus*, virulence, infection.

1. INTRODUÇÃO

1.1 *Staphylococcus saprophyticus* e Infecções de trato genito-urinário

Staphylococcus saprophyticus são bactérias Gram-positivas, coagulase negativas, novobiocina resistentes e frequentemente associada a infecções urinárias sem complicações. Staphylococci são considerados contaminantes da urina desde a década de 1960, quando foram isolados organismos coagulase-negativos possuindo antígeno 51 - referência a cepa 51877 isolada da urina de uma paciente com ITU aguda (TORRES PEREIRA, 1962). Este microrganismo foi classificado como micrococus do subgrupo 3 e posteriormente reclassificado como *S. saprophyticus* (RAZ; COLODNER; KUNIN, 2005).

Infecções do trato urinário (ITUs) são infecções comuns que podem ocorrer em qualquer ponto do trato urinário (uretra, bexiga, ureteres ou rins). Estas infecções são em sua maioria causadas por bactérias encontradas em tecidos próximos à abertura uretral que frequentemente colonizam a urina. Mulheres são geralmente mais suscetíveis a ITUs devido à proximidade entre a cavidade vaginal e reto em relação ao óstio externo da uretra (FOXMAN, 2014). As ITUs com manifestações sintomáticas mais comuns são a cistite – infecção da bexiga acompanhada pelos sintomas de disúria (dor ao urinar), dor na região supra-púbica, frequência urinária, urgência urinária e hematúria (presença de sangue na urina) – e pielonefrite - que é caracterizada pelos mesmos sintomas da cistite - acompanhados de febre, dor no flanco, sensibilidade na região costo-vertebral, náuseas e vômitos (FLORES-MEIRELES et al., 2015).

S. saprophyticus é a segunda maior causadora de ITUs em mulheres jovens sexualmente ativas, menos frequente que *Escherichia coli*, apenas. A prevalência de colonização de *S. saprophyticus* do trato urogenital em mulheres saudáveis é de 6,9% e o

lugar de colonização mais comum é o reto 40% (RUPP; SOPER; ARCHER, 1992). A incidência de infecções está sujeita a variações sazonais, em países temperados, ITUs por *S. saprophyticus* são mais comuns durante o verão e outono (BECKER; HEILMANN; PETERS, 2014). Um estudo conduzido no Brasil apontou uma maior incidência de ITUs por *S. saprophyticus* durante a primavera (Setembro a Dezembro) (LO et al., 2015). A infecção por *S. saprophyticus* pode ocorrer de várias formas como natação ao ar livre, contato com carnes contaminadas e principalmente via relação sexual (WIDERSTRÖM et al., 2012). O padrão utilizado para diagnóstico em análises urinárias são as fitas de urinálise, todavia estas não são sensíveis a detecção de organismos não redutores de nitrato, como bactérias Gram-positivas descartando-as como possíveis causas da infecção (KLINE; LEWIS, 2016). Desta forma, sugere-se que as infecções sub-sintomáticas causadas por bactérias Gram-positivas podem estar subestimadas por não serem detectadas por tiras de urinálise.

Ademais, staphylococci coagulase negativos são causadores de bacteremia nosocomial e infectam uma variedade de instrumentos prostéticos, principalmente em pacientes cateterizados, independente do gênero, devido a formação de biofilmes (ROGERS; FEY; RUPP, 2009; RUPP; ARCHER, 1994). Além de ser capaz de formar cálculos renais e uretrais compostos por estruvita (NH_4MgPO_4) devido a alcalinização da urina pela ação da enzima urease (FOWLER, 1985), é possível encontrar na literatura relatos de quadros mais graves resultantes de infecções por *S. saprophyticus* em humanos como óbito por bacteremia (HUR et al., 2016), endocardite resultando na substituição de valva do paciente (GARDUÑO et al., 2005), septicemia (GLIMAKER; GRANERT; KROOK, 1988) e abcesso cerebral assintomático apresentando hemorragia intracerebral (YOON; HA; KANG, 2019). *S. saprophyticus* não afeta apenas humanos e animais de

rebanho, tendo sido relatado como causador de cistite obstrutiva seguida de cistotomia em um gato persa (NIKOUSEFAT et al., 2018).

O genoma de *S. saprophyticus* (ATCC 15305) possui 2.516.575 pares de base com conteúdo G+C de 33,2%, similar ao descrito para outros staphylococci, contém 2446 ORFs, uma variedade de elementos móveis e dois plasmídeos. Curiosamente, *S. saprophyticus* apresenta abundância de ORFs envolvidos em transporte e regulação, enquanto apresentam uma baixa quantidade de ORFs relacionadas a elementos móveis e de virulência se comparado a outras espécies como *S. aureus* e *S. epidermidis*. Além disto, possui uma proteína ancorada a parede celular (hemaglutinina) e expansão paráloga única do sistema de transporte relacionado ao conteúdo iônico da urina, revelando grande especificidade para sobreviver neste ambiente (KURODA et al., 2005).

S. aureus, um dos maiores causadores de infecções nosocomiais, teve o genoma de duas cepas clínicas sequenciadas, uma cepa resistente a meticilina (N315) e à vancomicina (cepa Mu50) (KURODA et al., 2001). Apesar de *S. saprophyticus* e *S. aureus* serem do mesmo gênero e apresentarem certas características genômicas estruturais parecidas, a informação contida no DNA genômico diverge consideravelmente. O genoma de N315 possui 2.813.641 pares de bases e conteúdo G+C de 32,8%, enquanto Mu50 possui 2.878.084 e 32,9% de conteúdo G+C. As principais diferenças entre estas cepas são decorrentes da inserção de elementos móveis específicos de Mu50. Apesar da similaridade estrutural, *S. aureus* possui alta capacidade de adquirir genes de resistência por meio de transferência lateral de genes. A maioria dos genes relacionados a resistência a antibióticos estão relacionados a plasmídeos ou por elementos genéticos móveis incluindo uma ilha de resistência. Foram identificadas no genoma de *S. aureus* três ilhas de patogenicidade: síndrome de choque tóxico, exotoxinas e enterotoxinas. Nas últimas duas ilhas de patogenicidade, grupos de genes de exotoxina e

enterotoxina foram encontrados intimamente ligados a outros agrupamentos de genes que codificam fatores patogênicos putativos. Além destas características, *S. aureus* apresenta duplicações repetitivas de genes codificantes de superantígenos, sendo capaz de desencadear severas reações imunológicas em hospedeiros humanos (KURODA et al., 2001).

Após este primeiro sequenciamento completo do genoma de *S aureus*, outros estudos foram desenvolvidos para compreender o surgimento de cepas altamente resistentes principalmente à vancomicina.

1.2 Análises comparativas

Estudos proteômicos e genômicos comparativos tem sido conduzidos em algumas espécies do gênero *Staphylococcus* com o intuito de elucidar vias comuns e distintas em cepas com características fenotípicas diferentes.

A análise de proteômica comparativa entre uma cepa comensal (ATCC12228) e invasiva (ATCC35984) *S. epidermidis* revelaram que proteínas significativamente expressas estão envolvidas no transporte e metabolismo de carboidratos e degradação de lipídeos**.** A cepa ATCC35984 é capaz de formar biofilme enquanto a cepa ATCC12228 demonstra não possuir tal capacidade, todavia é importante salientar que mais de 60% dos isolados clínicos invasivos produzem biofilme mesmo na ausência do *operon ica*, responsável pela formação de biofilmes. Outra característica relevante é que a hemolisina é diferencialmente expressa entre as cepas. A hemólise induzida por toxina-σ foi observada após 24h de incubação em ATCC35984, em contrapartida a cepa ATCC12228 foi capaz de realizar hemólise apenas após 48h de incubação e em níveis consideravelmente baixos se comparada a cepa ATCC35984 (YANG et al., 2006). Estas

análises comparativas evidenciam o quanto, embora muito próximas geneticamente, as variações metabólicas são notórias mesmo entre cepas de uma mesma espécie e o quanto estas variações podem influenciar na patogenicidade de cada cepa.

Estudo conduzido com 236 cepas clínicas de *S. saprophyticus* traçou as principais variações genotípicas e fenotípicas presentes nesta espécie. O gene *capD* da cápsula de ATCC15305 foi detectado em 3 cepas apenas. Os genes regulatórios *agr*, *sarA*, e *rot* bem como o gene codificante de D-serina deaminase (*dsdA*), os fatores de uroaderência UafA e autolisina ligante a fibronectina Aas foram identificados em todas as cepas analisadas. Os demais fatores de virulência característicos de *S. saprophyticus* apresentaram distribuição heterogênea dentre as cepas analisadas. A enzima lipase associada a superfície Ssp esteve presente em 86,9% dos isolados clínicos, entretanto, apenas 66% das cepas mostraram possuir atividade lipolítica, sendo a cepa 7108 a cepa com maior atividade lipolítica detectada. A proteína ligante de repetição serina-aspartato de colágeno SdrI foi detectada em 10,2 % dos isolados clínicos, todavia 19% das cepas são capazes de aderirem a moléculas de colágeno e 18% das cepas negativas para SdrI aderiram ao colágeno, desta forma é muito possível que as cepas apresentem uma forma alternativa de se aderirem ao colágeno sem a necessidade de SdrI. Apesar de detectadas em todas as cepas, UafA e Aas apresentaram atividade de hemaglutinação em apenas 15,7% das cepas analisadas. Apenas 1,3% das cepas avaliadas eram acapsulares, sugerindo que a cápsula polissacarídica pode ter relevância para aumentar a chance de sucesso durante a infecção. (KLEINE; GATERMANN; SAKINC, 2010).

Recentemente, estudo exoproteômico comparativo foi realizado por nosso grupo de pesquisa para elucidar o perfil de proteínas secretadas por 3 cepas de *S. saprophyticus*. Neste estudo, as cepas ATCC 15305 (capsular), 7108 (acapsular) e 9325 (cápsula espessa) foram avaliadas quanto ao perfil de secreção de proteínas após cultivo em meio rico (BHI)

e foi possível detectar que diferenças quantitativas e qualitativas no perfil de expressão das 3 cepas. Foi possível identificar, por exemplo, maior secreção de proteases pela cepa acapasular 7108, enquanto as cepas ATCC 15305 e 9325 apresentaram secreção do fator de antigenicidade SsaA. A cepa 9325 apresentou a maior secreção de fatores antigênicos (OLIVEIRA et al., 2018). Os resultados sugerem que a diversidade de conteúdo de exoproteoma possa refletir na habilidade das bactérias desta espécie em utilizar maquinarias protéicas diferentes para promover a infecção. A **Tabela 1** apresenta as diferenças fenotípicas, genotípicas e de conteúdo de exoproteoma já caracterizadas para estas 3 cepas de *S. saprophyticus*, que também foram utilizadas no presente trabalho.

Tabela 1: Caracterização das cepas ATCC 15305, 7108 e 9325 de *S. saprophyticus*.

Característica	**ATCC 15305**	**7108**	**9325**
Presença de cápsula	Presente	Ausente	Presente, espessa
Tamanho da cápsula	Padrão	--	Espessa
uafA	Presente	Presente	Presente
dsdA	Presente	Presente	Presente
aas	Presente	Presente	Presente
sdrI	Ausente	Presente	Ausente
ssp	Ausente	Presente	Ausente
Secreção de SsaA	Presente	Ausente	Presente
Secreção de UroA	Presente	Presente	Ausente
Secreção de proteínas antigências	Presente, menor quantidade	Ausente	Presente, maior quantidade
Secreção de *moonlighting proteins*	Presente	Presente	Presente

Secreção de transportadores de sideróforos	Presente	Presente	Presente
Secreção de proteínas de resposta ao estresse	Presente	Presente	Presente

1.3 Fatores de virulência

1.3.1 Internalização de células

A internalização bacteriana por células não fagocíticas do hospedeiro é um mecanismo de virulência pois previne que as células bacterianas não sejam expelidas pela urina e possibilita a persistência da infecção. Tanto cepas capsulares (como por exemplo, ATCC15305) e acapsulares (por exemplo, 7108) apresentam capacidade de serem internalizadas por células de carcinoma da bexiga (linhagem 5637), sugerindo que a internalização não é dependente da cápsula (SZABADOS et al., 2008).

1.3.2 Urease

Vários patógenos bacterianos são capazes de utilizar ureia como fonte de nitrogênio através da atividade da enzima urease que converte ureia em amônia e ácido carbâmico. A hidrólise espontânea de ácido carbâmico a ácido carbônico resulta na produção de amônia.

Em condições fisiológicas, o próton do ácido carbônico se dissocia, as moléculas de amônia se tornam protonadas para formar amônia, ocasionando o aumento do pH no local da infecção, podendo interferir nas funções do hospedeiro (RUTHERFORD, 2014).

Urease é uma molécula heteropolimérica que possui múltiplos sítios ativos idênticos, cada um contendo dois íons Ni^{2+} ligados por uma lisina carbamilada e um íon hidróxido. A apo-urease é inativa e sua ativação demanda várias alterações químicas como hidrólise de GTP, carbamilação através do *uptake* de CO_2 e finalmente a transferência do íon de Ni^{2+} ao sítio ativo. Estas etapas são tipicamente catalisadas por quatro proteínas acessórias: UreD, UreF, UreG e UreE (MERLONI et al., 2014).

A urease tem sido descrita como fator de virulência em vários microrganismos. Em *Helicobacter pylori*, a urease desempenha papel crucial no estabelecimento da infecção. O silenciamento do gene codificante para a subunidade UreI em *H. pylori* resultou em um fenótipo incapaz de desempenhar a colonização gástrica e sobrevivência em pH < 4,0 (MOLLENHAUER-REKTORSCHEK et al., 2002). Para *Mycobacterium tuberculosis*, a atividade da urease é importante para a sobrevivência do patógeno em microambientes com limitação de nutrientes, na qual a ureia é sua única fonte de nitrogênio (LIN et al., 2012).

GATERMANN; JOHN; MARRE, (1989) foram os primeiros a descrever a urease como fator de virulência em *S. saprophyticus*, para tanto utilizaram a cepa 9325 e uma cepa mutante derivada de 9325 deficiente na produção de urease (GJ1187). Ao serem incubadas por 18h no meio de cultura com pH 6,5; a cepa 9325 deixou o meio básico (pH 8,55); enquanto a alteração no meio contendo a cepa GJ1187 foi insignificante (pH 6,80). Foram realizadas infecções em modelos animais com as cepas 9325 e a cepa mutante GJ1187 deficiente na produção de urease. A cepa 9325 apresentou maior persistências nos rins e bexiga, causou leucocitúria (eliminação de leucócitos na urina) elevada e em

15% dos animais infectados com esta cepa houve a formação de cálculos de estruvita ou carbonato de apatita na bexiga. Em avaliação macroscópica as bexigas dos animais infectados por 9325 estavam maiores e apresentaram paredes mais espessas. Em avaliações histológicas da bexiga, foi constatado que os animais infectados com 9325 tiveram o epitélio destruído com formação de abcessos, enquanto animais infectados com GJ1187 foi observado apenas o quadro de cistite moderada.

Devido seu papel durante a infecção, inibidores de urease tem sido estudados como agentes quimioterapêuticos no tratamento de ITU causadas por *S. saprophyticus*. A atividade de urease em extratos proteicos é suscetível a inibição de hidroxamatos, fosforodiamidatos e flavonoides, mas não a imidazois. Fluorofamida e ácido acetohidroxâmico pode atrasar temporariamente o aumento do pH causado por *S. saprophyticus*, mas não inibí-lo (LOES et al., 2014). A busca por métodos alternativos para o tratamento de ITUs é desejável a fim de evitar o possível surgimento de cepas resistentes a antibióticos e que causem o mínimo possível de efeitos colaterais.

1.3.3 Fatores de uro-aderência

A iniciação da doença requer contato íntimo do patógeno com o hospedeiro, as adesinas – complexos protéicos que reconhecem e se ligam a receptores também protéicos na superfície da célula do hospedeiro – são as moléculas responsáveis por esta função possibilitando a adesão da bactéria a superfícies abióticas ou ao tecido do hospedeiro. Em um ambiente *in vivo*, as adesinas facilitam a colonização mediando o primeiro e crucial contato com o tecido do hospedeiro. No caso de ITUs, a adesão não apenas protege as bactérias de serem eliminadas devido aos eventos de micção, como também pode mediar a invasão do tecido hospedeiro, através do qual a replicação pode ocorrer em um ambiente

intracelular protegido dos mecanismos de defesa do hospedeiro (WRIGHT; HULTGREN, 2006).

As proteínas ancoradas a parede celular, como as adesinas, possuem um conjunto comum de características que incluem um peptídeo sinal na região N-terminal, domínios funcionais que podem conter várias repetições em tandem ricas em serina e uma região hidrofóbica inserido em uma região da parede celular na porção C-terminal que contém o motivo LPXTG conservado e uma cauda carregada. O motivo LPXTG é necessário para a ancoragem da proteína na superfície celular e é clivada pela enzima sortase entre os resíduos de treonina e glicina durante a montagem de peptidoglicanos da parede celular (NAVARRE; SCHNEEWIND, 1999).

No genoma de *S. saprophyticus* foi detectada a presença da sequência codificante para o fator de uro-aderência A (*Uro-adherence factor A* – UafA) já descrito para outras espécies causadoras de ITUs desempenhando papel crucial no estabelecimento da infecção (KURODA et al., 2005). Este fator foi detectado em 76 isolados clínicos de três localidades diferentes (KING et al., 2011). O UafA de *S. saprophyticus* contém um peptídeo sinal na região N-terminal, uma região A de 72 kDa, seguido por um domínio B de 13 kDa que aparentemente é responsável pela adesão ao tecido hospedeiro, uma região R de 148 kDa rica em serina e glutamato e uma região abrangendo membrana citoplasmática e parede celular contendo o motivo LPXTG. A região A de UafA consiste em três subdomínios: N1, N2 e N3. Os domínios N2 e N3 apresentam 32% de similaridade em suas sequências. Os mecanismos de adesão e candidatos de ligação de UafA ainda não foram definidos, todavia esta moléculas demonstra desempenhar papel importante na atividade de hemaglutinação de *S. saprophyticus* (MATSUOKA et al., 2011).

Além de UafA, *S. saprophyticus* também contém o fator de uro-aderência B (UafB). O gene codificante para UafB encontra-se em um plasmídeo, flanqueado por uma sequência putativa de transposon, sugerindo a mobilidade deste elemento. Ao contrário de UafA, UafB apresenta uma distribuição muito mais restrita dentre cepas clínicas de *S. saprophyticus*, sendo detectado em apenas 4 de 76 cepas analisadas para a sequência.

A sequência de aminoácidos de UafB consiste em 2279 resíduos, contendo um peptídeo sinal atípico na extremidade N-terminal de 93 resíduos de comprimento. O segmento hidrofóbico abrangendo a parede celular na região C-terminal contém um motivo de sortase LPETG seguido por uma cauda carregada. UafB aparentemente se encontra ancorada a parede celular, mediando a ligação a células epiteliais da bexiga, fibronectina e fibrinogênio de humanos, interessantemente, esta glicoproteína não interage com células da bexiga de camundongos (KING et al., 2011), demonstrando alta especificidade quanto ao tipo de hospedeiro.

1.3.4 SdrI

SdrI (Serine-Aspartate repeat – sdr) é uma proteína que apresenta uma região contendo uma longa repetição de serina e aspartato de 854 aminoácidos. SdrI encontra-se ancorada a parede celular de cepas de *S. saprophyticus* e desempenha o papel de ligação a elementos da matriz celular de células do hospedeiro (SAKINC; KLEINE; GATERMANN, 2006). A princípio, quando descrita para *S. saprophyticus* cepa 7108, acreditava-se que esta proteína se ligava apenas ao colágeno (SAKINC; KLEINE; GATERMANN, 2006), todavia estudos posteriores revelaram que determinada região desta proteína possui um sítio de ligação capaz de se ligar também a fibronectina (SAKINÇ et al., 2009).

SdrI, assim como UafB, apresenta distribuição mais restrita dentre cepas de *S. saprophyticus*, sendo detectada em 7 de 123 cepas testadas para a presença do gene codificante para SdrI. A sequência genômica de SdrI codifica um polipetpídeo de 1893 aminoácidos, análises de homologia sugerem que este gene seria pertencente a família de proteínas de superfície Sdr. SdrI também contém o motivo LPXTG, típico de proteínas de superfície de organismos Gram-positivos (SAKINC; KLEINE; GATERMANN, 2006).

Testes *in vivo* de cepas mutantes de *S. saprophyticus* com deleção do gene codificante para SdrI demonstraram que as cepas são capazes de infectarem camundongos, todavia após 48h houve perda significativa da quantidade de bactérias na bexiga, sugerindo de SdrI não desempenharia um papel significativo no estabelecimento da infecção, entretanto é essencial para a persistência da infecção na bexiga e nos rins (KLINE et al., 2010).

1.3.5 Aas

Aas autolisina/adesina de *S. saprophyticus* possui 1463 resíduos aminoácidos em sua composição e o peso molecular de 159 kDa, contendo uma sequência de peptídeo sinal de 30 aminoácidos. Ao contrário das demais moléculas relacionadas a aderência ao tecido hospedeiro, Aas não apresenta o motivo LPXTG na região C-terminal. Todavia, a sequência do gene codificante para Aas é consideravelmente homólogo ao gene autolisina (*atl*) de *S. aureus* e ao gene *atl*E de *S. epidermidis*. Estas homologias estão presentes em quase toda a molécula, exceto pela região N-terminal da proteína. A Aas possui duas atividades enzimáticas distintas que são necessárias para a dispersão de agrupamentos durante a divisão celular e contém domínios de ligação a fibronectina e de atividade de hemaglutinina (HELL; MEYER; GATERMANN, 1998).

1.3.6 Ssp

Proteína associada a superfície de *S. saprophyticus* (*S. saprophyticus surface-associated protein* – Ssp) é um polipeptídeo inicialmente isolado de cepas 7108 de *S. saprophyticus* que produziam estruturas parecidas com “tufos de cabelo” quando cultivadas em membranas de diálise colocadas em meio de cultura de infuso de coração e cérebro sólido.

Esta proteína não causa hemaglutinação e quando expressa, Ssp forma projeções na superfície da célula no formato de fibras finas e curtas que tendem a formar feixes (GATERMANN et al., 1992). Estudos posteriores demonstraram que Ssp contém 755 aminoácidos e peso molecular de 83,5 kDa com peptídeo sinal de 37 aminoácidos e também apresenta alta homologia com lipases maduras de de staphylococci, tanto que quando cultivadas na presença de tributirina, cepas Ssp positivas mostraram atividade lipolítica. Cepas mutantes com deleção para gene codificante para Ssp não apresentaram atividade lipolítica.

Em contrapartida, como descrito para outros fatores de virulência de *S. saprophyticus*, Ssp não possui a capacidade de se ligar a elementos da matriz celular do hospedeiro, como o colágeno tipo I e também não apresenta o motivo LPXTG. De maneira geral, lipases são cogitadas como possíveis fatores de virulência em infecções como erupções e abcessos e estudos indicam que lipases são produzidas durante o processo infeccioso em camundongos (SAKINC et al., 2005), todavia seu papel na infecção em *S. saprophyticus* ainda não foi elucidado (KLINE; LEWIS, 2016).

1.3.7 SssF

Proteína F de superfície de *S. saprophyticus* (*S. saprophyticus surface protein F*) é uma proteína ancorada a parede celular de *S. saprophyticus* codificada pelo plasmídeo pSSAP2 da cepa MS1146 e altamente prevalente dentre isolados clínicos de *S. saprophyticus*. Esta proteína apresenta a sequência relativamente similar a proteína de superfície SasF de *S. aureus* que desempenha um importante papel na resistência ao ácido linoleico (KING et al., 2012).

O gene *sssF* codifica uma sequência de 654 aminoácidos com peso molecular de 73.5 kDa, peptídeo sinal de 45 resíduos e o motivo de ligação LPDTG na região C-terminal e um homólogo do gene *sssF* é codificado por pSSP1 da cepa ATCC15305. Esta proteína encontra-se amplamente distribuída dentre as cepas de *S. saprophyticus*, de 136 isolados clínicos analisados, 118 cepas foram SssF positivas (KING et al., 2012).

SssF não atua na mediação da adesão a células uroepiteliais ou colonização da bexiga por *S. saprophyticus*. Todavia, quando testadas para resistência ao ácido linoleico, cepas mutantes com deleção do gene *SssF* se mostraram menos viáveis em comparação a cepa selvagem e mutante reconstituído (KING et al., 2012). Desta forma, assim como a urease e ao contrário de fatores como UafA, UafB, Ssp e SdrI que são fatores importantes para a instalação e persistência da infecção, a SssF atua possibilitando uma maior viabilidade do patógeno no ambiente do hospedeiro.

1.4 Biofilmes

Biofilmes podem ser definidos como comunidade de microrganismos sésseis, compostos por células que estão aderidas a um substrato, interface umas às outras ou fixadas a uma matriz de substância extracelular polimérica e exibem fenótipo alterado

quanto ao crescimento, expressão gênica e produção de proteínas (DONLAN; COSTERTON, 2002). Estas estruturas começam a se desenvolver a partir da adesão reversível de microrganismos planctônicos a uma determinada superfície. Eventualmente, esta adesão se torna irreversível, possibilitando o desenvolvimento de microcolônias a partir destes microrganismos. A partir de então, os organismos passam a secretar substâncias poliméricas resultando na comunidade tridimensional madura do biofilme. Deste momento em diante pode ocorrer a dispersão de microrganismos da comunidade do biofilme em busca de novas superfícies a serem colonizadas (SENEVIRATNE, 2017). Biofilmes são ubíquos, estando presentes principalmente em ecossistemas aquáticos naturais, industriais e no ambiente hospitalar (COSTERTON et al., 1995).

Biofilmes foram descritos pela primeira vez por Zobell e Anderson (1936) ao perceberem uma correlação entre a densidade populacional de bactérias oriundas da água do mar, volume do meio e disponibilidade de superfície para adesão. Na indústria, biofilmes podem causar vários transtornos decorrentes de sua dispersão em sistemas hidráulicos fazendo-se necessário o uso de biocidas e testes em períodos regulares a fim de atestar a eficiência de eliminação de biofilmes (DONLAN; COSTERTON, 2002).

Várias infecções nosocomiais decorrentes ao uso de cateter venosos central, cateter urinário, valvas prostéticas e dispositivos ortopédicos estão relacionadas à biofilmes que se aderem à superfície de biomateriais (STEWART; COSTERTON, 2001). Na literatura médica, biofilmes foram tratados pela primeira vez por Marrie e colaboradores (1982) ao relatarem o caso de um paciente que fora admitido três vezes seguidas com quadro de bacteremia causado por *Staphylococcus aureus*. Ao investigarem a causa destas infecções, foi descoberto que o eletrodo do marcapasso do paciente continha biofilmes de *S. aureus*. Após a substituição do eletrodo contaminado, não houve reincidências. Com o avanço em técnicas e novas tecnologias aplicadas à medicina

atrelado ao aumento da média geral da expectativa de vida em muitos países, o uso de implantes com finalidade estética e dispositivos médicos tem aumentado consideravelmente. Em 2004, um levantamento a respeito do impacto de infecções bacterianas relacionada ao uso de implantes no Estados Unidos estimou taxas de infecção de 2% para próteses articulares e implantes de mama, 4% para próteses valvares mecânicas, marcapassos e desfibriladores, 10% para válvulas ventriculares e 40% para dispositivos de assistência ventricular (DAROUICHE, 2004).

Mesmo em indivíduos imunocompetentes, infecções causadas por biofilmes são dificilmente eliminadas. O uso de antibióticos pode sanar os sintomas de infecção por eliminar as bactérias em suspensão, todavia é ineficaz no combate aos microrganismos contidos no biofilme (STEWART; COSTERTON, 2001). *S.* aureus, por exemplo, chegam ter a suscetibilidade a antibióticos reduzida de 100 a 1000 vezes em biofilmes (CERI et al., 1999).

Biofilmes apresentam grande heterogeneidade química, estrutural e biológica. As células que compõem o biofilme não são apenas distintas das células planctônicas, mas também são distintas entre si tanto espacialmente como temporalmente ao longo do processo de desenvolvimento do biofilme (STEWART; FRANKLIN, 2008). Mesmo o biofilme derivado de uma célula apenas pode conter uma população espectro fisiológico variado. Sendo células que crescem rapidamente, células que crescem mais lentamente, células em estado parecido com a fase estacionária ou completamente dormentes. Podem também estar presentes células crescendo de maneira aeróbia ou fermentativa (DONLAN; COSTERTON, 2002). Esta diversidade é vital para a persistência destes microrganismos, pois os organismos metabolicamente mais ativos mantém os que se encontram em estado dormente. Em compensação, caso haja a eliminação em larga escala

destes organismos mais ativos, os que estavam em estado dormente serão responsáveis pela recolonização do ambiente (DONLAN; COSTERTON, 2002).

Durante o processo de formação do biofilme, surgem variações fenotípicas e genotípicas que asseguram maior sucesso na colonização de superfícies. Em *Pseudomonas aeruginosa*, o fenótipo rugoso produz um biofilme de matriz mais resistente e garante maior resistência a antibióticos, estresse oxidativo. Este fenótipo surge longo do processo de desenvolvimento do biofilme e apresenta caráter hereditário, indicando que estas alterações são de natureza genética. (BOLES; THOENDEL; SINGH, 2004). Por outro lado, células adjacentes no biofilme podem apresentar perfis de expressão diferentes para o mesmo gene mesmo estando estas células sob as mesmas condições de (EASTBURN et al., 2002; SERRA et al., 2013) sem necessariamente passar por eventos de alteração genética.

A substância polimérica extracelular (SPE) é composta majoritariamente por polissacarídeos, proteínas, ácidos nucleicos e lipídeos, promovendo a estabilidade mecânica, mediando sua adesão à superfícies e formar uma estrutura tridimensional coesa que interconecta e imobiliza temporariamente células do biofilme (FLEMMING; WINGENDER, 2010). A obtenção de nutrientes se dá pela secreção de substâncias como compostos captadores de nutrientes e enzimas digestórias (NADELL et al., 2015) e redes intricadas de canais fluem através destas estruturas promovendo a acessibilidade de nutrientes mesmo nas regiões mais profundas dos biofilmes (DONLAN; COSTERTON, 2002).

1.4.1 Biofilmes em Staphylococci

Em *S. epidermidis* e *S. aureus*, o locus *icaADBC* está envolvido na síntese de polissacarídeo de adesão intercelular (PAI) localizada principalmente na superfície celular. PAI apresenta estrutura linear composta por β-1,6- glicosaminoglicanas de pelo menos 130 2-deoxi-2-amino-d-glicopiranosil dos quais 80-85% são N-acetiladas; o restante não acetilado é positivamente carregado (HEILMANN et al., 1996). A ausência deste locus afeta diretamente a capacidade de *S. epidermidis* e *S. aureus* (CRAMTON et al., 1999) de produzirem biofilmes.

O produto do gene *icaA* é uma proteína transmembranar com homologia a N-acetil-glicosaminiltransferases e necessita do produto do gene *icaD* para atividade ótima. Oligômeros de N-acetil-glicosamina produzidos por icaAD atingem o comprimento máximo de 20 resíduos e apenas quando *icaAD* é coexpresso com *icaC*, o qual codifica uma proteína putativa de membrana, que cadeias mais longas de oligômeros são sintetizadas. O produto gênico de *icaB* é responsável pela desacetilação de poli-*N*-acetilglicosamina. Quando a desacetilação não ocorre, as poli-acetil-glicosaminoglicanas tem sua aderência a superfície bacteriana e mediação no desenvolvimento de biofilmes são comprometidos (CUE; LEI; LEE, 2012). A regulação do operon se dá por meio do gene *icaR* que se encontra na região *upstream* ao operon icaADBC. O gene codifica um fator de transcrição da família TetR que aumenta a expressão do operon em condições de estresse osmótico (NaCl) ou na presença de etanol (CONLON; HUMPHREYS; O'GARA, 2002).

A adesão inicial de staphylococci é geralmente mediado por proteínas da superfície celular que se liga à matriz extracelular/plasma das células do hospedeiro como fibrinogênio, fibronectina, colágeno, vitronectina ou laminina (CUE; LEI; LEE, 2012).

Cepas de *S. aureus* também podem dispor de mecanismos independentes do operon *ica* para a síntese de biofilmes. Proteínas da família BAP (*Biofilm Associated Protein* – Proteína associada ao biofilme) atuam na adesão intercelular, acumulação de cluster celulares em camadas e, assim como *ica*, na adesão primária em superfícies abióticas (CUCARELLA et al., 2001).

1.5 Síntese de nucleotídeos

Nucleotídeos desempenham funções diversas no contexto celular. Além de serem os monômeros que compõem os ácidos nucleicos, atuam como carreadores essenciais de energia química, cofatores e segundos-mensageiros celulares. Estas moléculas podem ser sintetizadas por vias de salvação e *de novo*. As vias de salvação reciclam as bases livres e os nucleosídeos liberados a partir da degradação de ácidos nucleicos. A síntese *de novo* parte dos precursores metabólicos como aminoácidos, ribose-5-fosfato, CO_2 e NH_2 (NELSON; COX, 2014)

A síntese de purinas é mais complexa que a de pirimidinas. A glutamina doa o radical -NH_2 para fosforribosil pirofosfato (PRPP) antes de a inosina 5'- monofosfato (IMP) ser sintetizada com a adição de carbonos e nitrogênios nas formas de glicina, metenil tetrahidrofolato, glutamina, aspartato e formil-tetrahidrofolato. IMP é convertido a adenosina 5'-monofosfato (AMP) e guanosina 5'-monofosfato (GMP). A PRPP amidotransferase, que catalisa a primeira etapa da via, é regulada por *feedback* negativo pelos produtos finais da via, AMP e GMP. AMP e GMP são fosforilados a dinucleotídeos em reações catalisadas por nucleotídeo-quinases que consomem ATPs. GDP é fosforilada a GTP em uma reação similar, ADP é convertida a ATP pelos mecanismos convencionais de fosforilação a nível de substrato ou pela ATPase ligada à membrana (KIM; GADD, 2008).

A via de síntese de purinas está organizada em um operon nos genomas de staphylococci. Este sistema de regulação gênica é comum em procariotos, sendo que cerca de 38% do genoma de *S. aureus* encontra-se organizado na forma de operons (WANG et al., 2004).

Escassez de precursores de nucleotídeos são essenciais para a limitação do crescimento bacteriano (SAMANT et al., 2008) e a deleção de genes do operon da síntese *de novo* de purinas reflete de maneira variada sobre o metabolismo bacteriano. Defeitos no metabolismo de purina podem deixar bactérias mais suscetíveis aos ataques oxidativos em uma condição, provavelmente devido a incapacidade de reparar os danos causados pelas espécies reativas de oxigênio do sistema imune do hospedeiro (MANTENA et al., 2008). A eliminação da via de síntese *de novo* em UPEC (*Uropathogenic Escherichia coli – E. coli* uropatogênica) não afeta propriedades de adesão *in vitro* e *in vivo*, todavia sua capacidade de expansão comprometida após internalização por células epiteliais de bexiga (SHAFFER et al., 2017). Em contrapartida, a inativação do repressor transcricional da biossíntese de purinas *purR* aumenta o potencial patogênico de *S. aureus* em decorrência da superexpressão de proteínas ligantes a fibronectinas. Anticorpos contra estas proteínas protegem o hospedeiro da hipervirulência causada pela ausência de *purR* (GONCHEVA et al., 2018; MCLEAN et al., 2019).

OBJETIVOS

Os objetivos do presente trabalho são:

- Descrever o proteoma global de três cepas patogênicas de *S. saprophyticus* com características fenotípicas e genotípicas diferentes;
- Comparar o perfil proteômico das cepas avaliadas;
- Validar os dados proteômicos obtidos;
- Identificar estratégias de virulência diferenciais entre as cepas;
- Avaliar a importância da via de produção de purinas de *S. saprophyticus* na formação de biofilme;
- Propor potenciais proteínas de virulência como alvos para desenho racional de fármacos;

MANUSCRITO I

Staphylococcus saprophyticus proteomic analyses elucidate differences in the protein repertories among clinical strains related to virulence and persistence

Karla Christina Sousa Silva[1], Lana O'Hara Souza Silva[1], Guilherme Algusto Alves Silva[1], Clayton Luiz Borges[1], Evandro Novaes[2], Juliano Domiraci Paccez[1], Wagner Fontes[3], Marcia Giambiagi-de-Marval[4], Célia Maria de Almeida Soares[1], Juliana Alves Parente-Rocha[1]*

[1]*Laboratório de Biologia Molecular, Instituto de Ciências Biológicas, Universidade Federal de Goiás, Goiânia, Goiás, Brazil.*

[2]*Departamento de Biologia, Universidade Federal de Lavras, Lavras, Minas Gerais, Brazil.*

[3]*Laboratório de Química de Proteínas, Instituto de Biologia, Universidade de Brasília, Brasília, Distrito Federal, Brazil.*

[4]*Laboratório de Microbiologia Molecular, Instituto de Microbiologia Prof. Paulo de Góes, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Rio de Janeiro, Brazil.*

KCSS: karlabio@live.com

LOHSS: lanaohara.loss@gmail.com

GAAS: algustoguilherme1@gmail.com

CLB: clayton@ufg.br

EV: evnovaes@gmail.com

WF: wagnerf@unb.br

MGM: marciagm@micro.ufrj.br

CMAS: cmasoares@gmail.com

JAPR: juparente@ufg.br

* Corresponding author: Juliana Alves Parente-Rocha. Universidade Federal de Goiás. Avenida Esperança, s/n ICB2, Laboratório de Biologia Molecular. 74690-900.

Goiânia, GO, Brazil. juparente@ufg.br.

Abstract

Staphylococcus saprophyticus is a gram positive and coagulase negative cocci that composes the skin microbiota and can act as opportunistic agent causing urinary tract infections, being more frequent in sexually active young women. The ability of a pathogen to cause infection in the host is associated to its ability to adhere to host cells andto survive to host immune defenses persisting inner the host. In this work, we presented the comparative proteomic profile of three *S. saprophyticus* strains. It was possible to characterize differences in the proteome content, specially related to expression of virulence factors related to pathogenicity. We compiled this data and previous data - concerning to phenotypic characteristics and exoproteome profile of these strains - and we detected one strains (9325) possessing higher production and secretion of proteins related to virulence. Our results show that phenotypic, genotypic and proteomic differences reflect in the ability to survive during interaction with host cells, since 9325 strain presented higher survival rate after macrophage interaction. In counterpart, the 7108 strain that possesses lower content of proteins related to virulence presented higher ability to form biofilm suggesting that this strain can be better adapted to persist in the host and in the environment. Our work describes, for the first time, proteomic flexibility among *S. saprophyticus* strains reflecting in virulence and persistence.

Key words: *Staphylococcus saprophyticus*; proteome; virulence; urease; thioredoxin; biofilm.

INTRODUCTION

The gram-positive and coagulase negative cocci *Staphylococcus saprophyticus* is the causative agent of urinary tract infections, being more frequently causing infection in sexually active young women [1]. The periurethral region can be a repository of this species, as well as, skin and mucosal regions [2]. *S. saprophyticus* can compose the skin microbiota and act as opportunist bacteria. Studies with contact sports athletes demonstrates high *S. saprophyticus* prevalence in the skin, suggesting contact among athletes can function as bacteria spreader [10]. *S. saprophyticus* can possess polysaccharide capsule that enhance virulence in animal model [10] but does not enhance internalization rate into human bladder cells [11]. In addition, non-capsular strains have been less frequently isolated from clinical strains (around 1.3%) [12], suggesting the capsule is not required to cause infection.

The ability of *S. saprophyticus* to cause infection can be attributed to virulence factors, such as urease [3, 4], surface proteins [5] and D-serine-deaminase protein (DsdA) [6]. Comparative analysis using genomic approach describes that *S. saprophyticus*, compared to other coagulase negative *Staphylococcus* species, lacks many of the adhesion proteins and other virulence factors that can explain differences at a clinical level [7]. In counterpart, *S. saprophyticus* presents importance to the public health not only by ability to cause human infections, but also by the ability to persist in the environment and acquire and transmit plasmids that can confer antibiotic resistances [8, 9].

Staphylococcus species can form biofilms dependent on polysaccharide intercellular adhesin (PIA), synthesized by *ica* operon [10]. Biofilms can reduce access of the host defense system to *Staphylococcus* and impair antibiotic action. Also, conjugation can occurs at higher levels in staphylococcal cells in biofilms compared to planktonic cells [11]. Analysis of 169 *S. saprophyticus* strains shows that 70% of these

strains possess ability to form biofilm. In addition, the biofilm formation increases the resistance to five antibiotics (vancomycin, oxacillin, trimethoprim/sulfamethoxazole, ciprofloxacin, and norfloxacin) in around 32 folds [12].

Genotypic variation among *S. saprophyticus* strains has already been described. Studies performed with 236 *S. saprophyticus* strains obtained from patients demonstrated that 100% of the strains possess genes encoding virulence factor, such as, urease, DsdA, uroadherence factor a (UafA) and autoloysins (Aas) suggesting that these genes are required for infection. In counterpart, gene encoding the surface protein SdrI is detected only in 10% of the strains, suggesting it is not an essential gene for infection [13]. A Previou work by our research group analyzed comparative proteomic data among capsular and non-capsular *S. saprophyticus* strains elucidated the *S. saprophyticus* exoproteomic repertoire and showed that different strains possess different secreted machinery that can be used during infection. For example, the highly capsular 9325 strains secretes higher content of antigenic proteins and transglycosylases while the non-capsular 7108 strain does not secrete the SsaA antigenic protein and secretes higher content of proteases. These results show diversity in protein secretion among strains [14].

In our work, we used a proteomic approach to compare three *S. saprophyticus* strains possessing different patterns of capsule: the ATCC 15305 possessing capsule, the non-capsular 7108 strain and the highly capsular strain 9325. Of special note, ATCC 15305 produces higher content of proteins for purine biosynthesis. On the other hand, 9325 secretes higher level of urease. The production of thioredoxin, related to oxidative stress response, is also different among the strains. Our results show that these strains use different molecular machineries that can confer ability to cause infection. The ability to produce, secrete and use virulence factors is deeply related to the bacterial survival during

infection and each strain can use a different repertoire, related to its metabolic flexibility to cause infection in the host.

MATERIAL AND METHODS

S. saprophyticus strain maintenance and culture conditions

S. saprophyticus clinical strains ATCC 15305, 7108 and 9325, were used in this study. These strains were previously phenotypically characterized and they differ in the presence of capsule (ATCC 15305 possess capsule, 7108 is non-capsular and 9325 possess a thick capsule) in the presence of virulence factors and in the exoproteome content [13-15]. *S. saprophyticus* cells were cultured in BHI medium (Sigma-Aldrich, St. Louis, MO, USA) and stored at -80º C in 50% (v/v) glycerol. In order to obtain the protein extracts, a single *S. saprophyticus* colony of each strain was pre-incubated separately in BHI medium until the stationary phase (after around 20 hours) with shaking at 36º C. After, the inoculum was performed using 1% of the pre-inoculum and the cells were incubated at 36º C with shaking until the optical density 0.6 at 600 nm wavelength using spectrophotometry SpectraMax Paradigm (Molecular Devices, Lagerhausstrasse, Austria).

Cell growth curve of *S. saprophyticus* strains

Each strain was cultured for the pre-inoculum and inoculum, as described above. The inoculum was split into 200 μl aliquots and placed in 96-well plates and left under agitation at 37° C. Cell growth was monitored each 2 h for a 10 h period through spectrophotometry using SpectraMax Paradigm (Molecular Devices, Lagerhausstrasse, Austria).

Obtaining the proteins extracts from *S. saprophyticus*

Protein extracts were obtained by using glass beads 0.2 to 0.8nm (Sigma Aldrich, St Louis, MO, USA) to disrupt the cells using Bead Beater (Biospec, Bartlesville, OK,

USA) during three cycles of 30 seconds. After, the protein extract was obtained by centrifugation at 10000 *g* for 15 minutes. The protein extracts were quantified using Bradford reagent (Sigma Aldrich, St Louis, MO, USA) using spectrophotometry SpectraMax Paradigm (Molecular Devices, Lagerhausstrasse, Austria). The integrity of the protein extract obtained was evaluated on SDS-PAGE containing 30 µg of extract of each sample. Biological triplicates were obtained and used for proteomic analysis.

Digestion of protein extracts for nano-ESI-UPLC-MS^E acquisition

A total of 150 µg of each protein extract was trypsin digested. The protocol used for digestion was previously described [14]A total of 20 ng of trypsin (Promega, Madison, WI, USA) was used for each sample. The samples were then treated with trifluoroacetic acid (TFA) 0.1% (Sigma Aldrich, St Louis, MO, USA) and purified on C18 ZipTips (Millipore, Darmstadt, Germany). The resins were equilibrated with acetonitrile (ACN) prior to use (Sigma Aldrich, St Louis, MO, USA), followed by washing with the following solutions: (1) 80% ACN and TFA 0.1% (2) 50% ACN and 0.1% TFA, (3) 30% ACN and 0.1% TFA and finally (4) 0.1% TFA.

Data processing and protein identification

The obtained raw mass spectrometry data were processed using the ProteinLynx Global Server v3.0.2 (PLGS) (Waters Corporation, Milford, MA, USA). For protein identification, processed spectra were searched against *S. saprophyticus* protein sequences (from Uniprot Proteomes) together with reverse sequences for false positive rate (FDR) estimation.. The mass error tolerance for peptide identification was under 50 ppm. Criteria for protein identification included: (i) at least 2 fragment ions per peptide, (ii) at least 5 fragment ions per protein, (iii) at least 1 peptide per protein, (iv) carbamidomethylation of cysteine as a fixed modification, (v) phosphorylation of serine, threonine and tyrosine; and methionine oxidation as variable modifications, (vi)

maximum protein mass of 600 kDa, (vii) 1 missed cleavage site was allowed for trypsin, (viii) maximum of 5% of FDR was allowed. The protein and peptides table generated by PLGS were merged and the Dynamic Range, Peptide Detection Type and Mass accuracy were determined for each sample using MassPivot v3.1, FBAT and Spotfire® v7.0.0, program (TIBCO software), as previously described [23]. Microsoft Excel® (Microsoft) was used for table manipulations.

Statistical analysis of proteomic data

The statistical analysis was conducted using data from biological and experimental replicates. The amount of each protein, from each *S. saprophyticus* strain and replicate, was used as a measure of protein expression, as described previously [14]. The R software was used and the expression data were $\log_2$ transformed and quantile normalized with the limma package [16] using the normalize BetweenArrays function. Differential abundance analyses between the three *S. saprophyticus* strains were performed with an empirical Bayes method implemented in the limma package [17]. Proteins were declared differentially abundant using a threshold of 0.05 false discovery. The functional classification was performed by using Uniprot (http://www.uniprot.org).

Urease activity assays

The urease activity was assayed as described [24]. PEB buffer (100 mM Sodium phosphate; 10 mM EDTA) was added to the samples containing 10 μg of protein extract and 50 mM urea, heated to 37° C for 30 minutes. Then 80 μl alkaline hypochlorite and 80 μl phenol nitroprusside solution were added to the samples and heated to 50° C for 6 minutes. The absorbance was read at 625 nm using SpectraMax Paradigm (Molecular Devices, Lagerhausstrasse, Austria). The readings were compared to the standard curve with NH_4Cl concentrations ranging from 0 μM to 500

μM (0; 5; 10; 25; 50; 100; 150; 200; 250; 300; 500 μM NH_4Cl).

Bacterial cells were also cultured on urea agar-based plates, as described by (Christensen, 1946). Cells were plated in different dilutions (10^2; 10^3; 10^4; 10^5 cells) and kept for 20h at 37° C.

Thioredoxin assay

A total of 10^9 bacterial cells – counted in cytometry flow - from each strain were lysed with 0,5 ml Lysis Buffer (50 mM Tris-HCl; 150 mM NaCl; 50 mM EDTA; pH 7,2) and approximately 0,5 g of glass beads. The samples were vortexed for 3 cycles of 3 minutes with 1 minute interval. Cellular debris were removed by centrifugation. 100 μl of supernatant plus 100 μl 500 mM phosphate buffer (pH 7,5) and 1 mM DTNB (5,5 dithio-bis-(2-nitrobenzoic acid) were added to each sample. The reaction took place at room temperature for 15 minutes. The absorbance was read at 412 nm using SpectraMax Paradigm (Molecular Devices, Lagerhausstrasse, Austria).

***S. saprophyticus* interaction assay with macrophage cells**

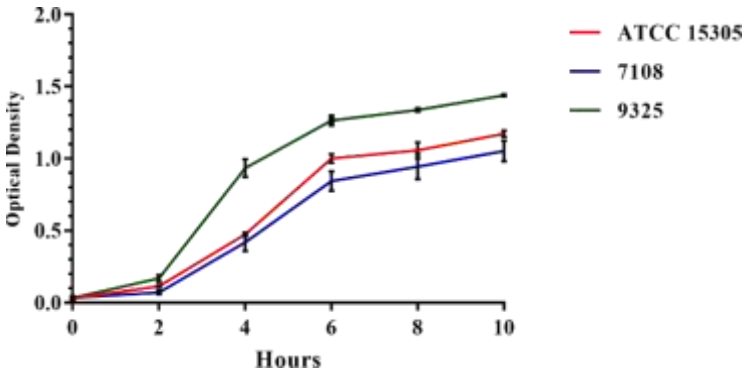
The interaction assay was performed as previously described [25]. Macrophages J774 1.6 derived from *Mus musculus* (Banco Central do Rio de Janeiro, Rio de Janeiro, RJ, Brazil) were cultured in RPMI medium with 1% amino acid solution and 10% Fetal Bovine Serum. A total of 10^6 macrophages were placed into 6-wells plates and activated with Interferon-γ 24h prior to infection to a final concentration of 1 unit/ml. Bacterial cells were cultured in BHI broth to the optical density of 0.2 at 620 nm, harvested through centrifugation (3000 rpm for 5 minutes), washed with saline solution 0.9 % and resuspended in RPMI medium without antibiotics. The macrophage medium was replaced with the medium containing $50x10^6$ *S. saprophyticus* cells. After 2h infection, 50 μl of supernatant of each well was plated on BHI agar plates in triplicates for each well. Each well was washed three times with 0,9% saline solution and the macrophages lysed for 10 minutes with 1.5 ml of ice-cold ultrapure water. After lysis, bacterial cells

were centrifuged and resuspended in 300 µl 0.9% saline solution and 50 µl plated onto BHI agar plates in triplicates for each well. The plates were kept for 20h at 37° C and colony formation units (CFU) counted for analysis.

RESULTS

Evaluation of *S. saprophyticus* strains cell growth

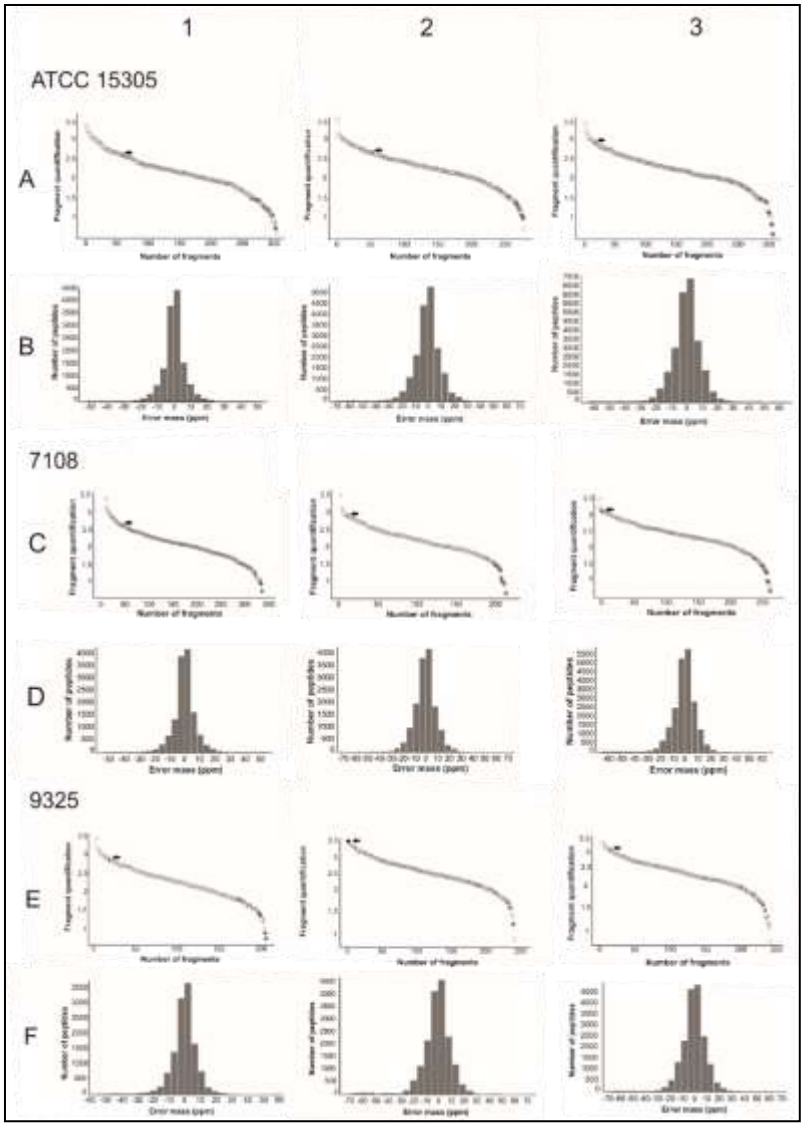
In order to perform the comparative proteomic analyses using cell from the *S. saprophyticus* strains at the same point of the cell growth curve, we performed the evaluation of cell growth using spectrophotometer, as described above. The result is shown in the Supplementary Figure 1. The result shows that all the strains present the same profile of cell growth. However, since 9325 strain presented higher values of optical density we evaluated the cell growth using flow cytometry method and the number of cells in each time point analyzed was similar among the strains (data not shown). We speculate that the capsule size can interfere in the optical density and can explain the differences shown in the Figure 1.



Supplementary Figure 1 – *S. saprophyticus* cell growth curve. The graphic demonstrates *S. saprophyticus* strains growth behavior in 10h period.

Proteomic profile of *S. saprophyticus* strains

Protein extracts were obtained from three biological replicates. Trypsin digested protein extracts were quantified by Nano-UPLCMS^{E} and protein and peptide data were generated by PLGS. A total of 276 proteins were detected in this work. The description of the identified proteins is shown in the Supplementary Table 1. To estimate false positive rates, a reverse sequence database was created and used in protein identification. This strategy resulted in replicate 1, 2 and 3 rates of 1.86%, 0% and 0.46% for ATCC 15305, 0.89%, 2.38% and 0.57% for 7108; and 1.94%, 0.70% and 0.70% for 9325, respectively. The number of identified peptides was 14,384; 18,995 and 26,328 in ATCC 15305; 14,244; 15,438 and 21,854 in 7108; and 12,890; 17,785 and 19,055 in 9325 in replicate 1, 2 and 3, respectively. Regarding peptides parts per million errors (ppm), the majority (77.8%, 76.0% and 75.0% for ATCC 15305; 78.5%, 77.4% and 76.1% for 7108; and 77.4%, 74.7% and 75.9% for 9325 in replicate 1, 2 and 3, respectively) was detected with an error of less than 10 ppm. and Dynamic Range detected a 3-log range concentration and a good distribution of high and low molecular weights in all samples (Supplementary Figure 2). The identified proteins were classified and the functional categorization is shown in Figure 1.



Supplementary Figure 2 – Quality analysis of proteomic data from *S. saprophyticus* ATCC 15305, 7108 and 9325. 1,2 and 3 represents replicates 1, 2 and 3 respectively. Detection of dynamic range of proteomic analysis. Protein data were used to calculate the order of magnitude of protein identification. Regular, reverse and standard proteins are represented in blank, grey and black (arrow), respectively. Regular and reverse proteins indicate identified proteins using regular and reverse proteome of *S. saprophyticus*. (A) ATCC 15305, (C) 7108 and (E) 9325. Peptide Mass accuracy analyses. Peptide data were used to make the bar graph showing the accuracy of mass for peptides. (B) ATCC 15305, (D) 7108 and (F) 9325.

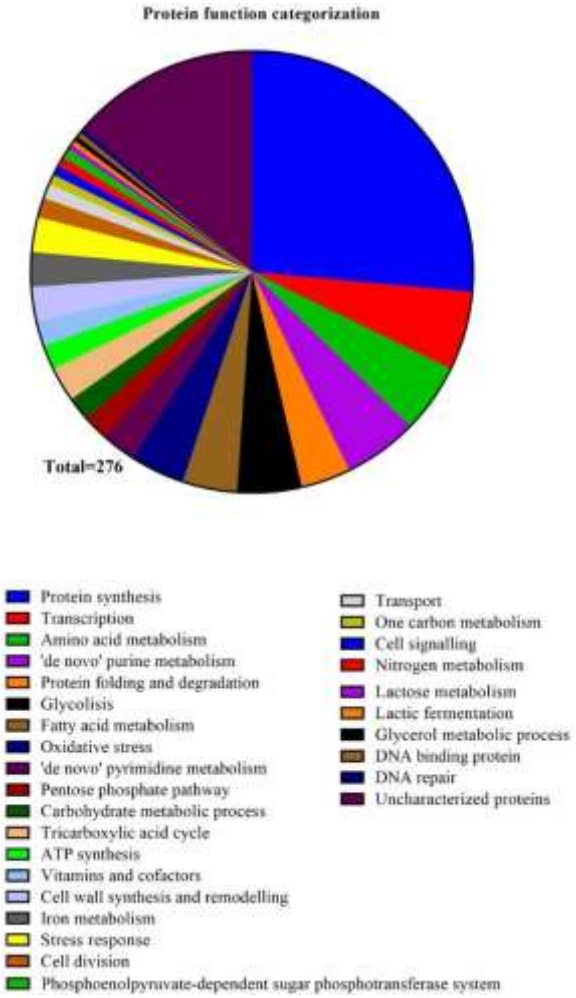


Figure 1 – Protein functional categorization. Total of proteins identified through mass spectrometry and functionally classified according to Uniprot database (www.uniprot.com).

Among the non-regulated proteins, identified in the three *S. saprophyticus* strains, it was possible to detect proteins related to transcription and translational processes, glycolysis pathway and amino acid metabolism. A total of 170 proteins were identified as non-regulated among the strains (Supplementary Table 1).

Differentially expressed proteins among *S. saprophyticus* strains

In order to detect differentially expressed proteins among the three *S. saprophyticus* strains, we analyzed the proteins presenting statistical significantly

expression values among them. It was possible to detect proteins related to virulence and host pathogen interaction differentially regulated among the strains. Of special note, we selected proteins related to oxidative stress, urease system and *de novo* purine biosynthetic process. These proteins are listed in the Table 1.

Table 1 - Proteins related to virulence differentially expressed among S. saprophyticus strains.*

Accession number[1]	Protein description	Log FC[2]			p-value[3]		
		ATCC* vs 7108	ATCC* vs 9325	7108 vs 9325	ATCC* vs 7108	ATCC* vs 9325	7108 vs 9325
Oxidative stress							
Q49WR2	Thioredoxin	0.494	-0.133	-0.627	0.011	0.480	0.001
Q49YE4	Probable thiol peroxidase	0.280	-0.419	-0.699	0.376	0.085	0.005
Q49UT8	Alkyl hydroperoxide reductase subunit C	0.490	-0.408	-0.899	0.026	0.040	0.000
Q49XC1	Catalase	3.691	3.811	0.120	0.000	0.000	0.735
Q49XN4	Peptide methionine sulfoxide reductase MsrB	0.228	0.460	0.232	0.449	0.044	0.339
Q49XZ6	Superoxide dismutase [Mn/Fe]	0.338	-0.406	-0.744	0.128	0.038	0.000
Q49YE0	Putative universal stress protein SSP1056	2.055	2.055	0.000	0.000	0.000	1.000
Q49UU5	Nitronate monooxygenase	-0.339	1.063	1.402	0.402	0.001	0.000
Nitrogen metabolism							
Q4A0J5	Urease subunit alpha	-1.782	1.338	3.120	0.034	0.068	0.000
Q4A0J8	Urease accessory protein UreG	-0.416	0.936	1.352	0.464	0.038	0.003
***de novo* purine biosynthetic pathway**							

Q49WI9	N5-carboxyaminoimidazole ribonucleotide synthase PurK	2.377	2.377	0.000	0.000	0.000	1.000
Q49WJ0	Phosphoribosylaminoimidazole-succinocarboxamide synthase PurC	2.184	2.184	0.000	0.000	0.000	1.000
Q49WJ1	Phosphoribosylformylglycinamidine synthase subunit PurS	1.035	1.035	0.000	0.000	0.000	1.000
Q49WJ2	Phosphoribosylformylglycinamidine synthase subunit PurQ	1.006	1.006	0.000	0.011	0.007	1.000
Q49WJ3	Phosphoribosylformylglycinamidine synthase subunit PurL	3,345	3,345	0,000	0,000	0,000	1,000
Q49WJ5	Phosphoribosylformylglycinamidine cyclo-ligase PurM	2,365	2,365	0,000	0,000	0,000	1,000
Q49WJ7	Bifunctional purine biosynthesis protein PurH	4,249	4,249	0,000	0,000	0,000	1,000
Q49WJ8	Phosphoribosylamine--glycine ligase PurD	3,038	3,038	0,000	0,000	0,000	1,000

[1] Accession number provided by Uniprot Database (http://www.uniprot.org/).

[2] Obtained from limma's topTable by subtracting the average expression in log2 scale against the strains.

[3] Proteins with p-value ≤ 0.05 were considered regulated among the strains. p-value from the Student's t distribution.

* ATCC 15305 strain.

We detected thioredoxins lower expressed in the 7108 strains when compared to ATCC 15305 and 9325 strains. The expression of thioredoxin was not statistically different when comparing the ATCC 15305 and 9325 strains. In order to validate the proteomic data we performed the thioredoxin assay, as described above. The result is shown in the Figure 2. The result of the enzymatic assay corroborates our proteomic analysis, showing a lower level of thioredoxin expression in the strain 7108.

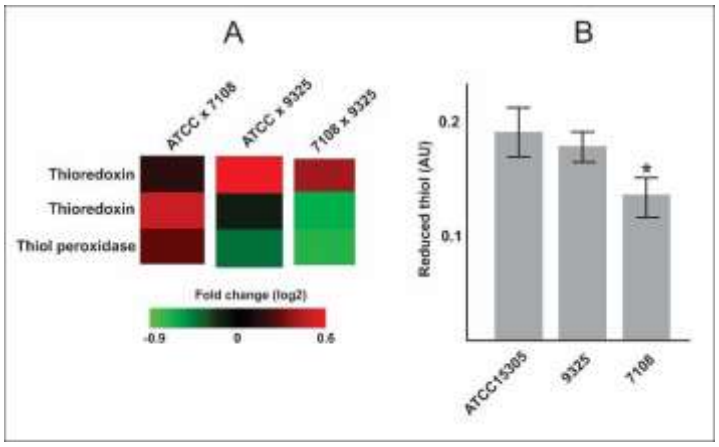


Figure 2 – Thioredoxins and thiol peroxidases expression and enzymatic assay. (A) Heat map of protein expression. Heat map showing fold change (log2) comparing thioredoxins and thiol peroxidase expression among *S. saprophyticus* strains. The heat map was generated using proteomic data. **(B) Enzymatic assay of thiol reduction.** The reactions were performed using protein extracts from the three *S. saprophyticus* strains. Reduced thiol formed was measured. The assay was performed in biological duplicate and experimental triplicate. Asterisk indicates statistical significance using Student's T test.

The 7108 *S. saprophyticus* strain presented a very expressive amount of two subunits from urease system, compared to the strains 9325 and ATCC 15305. The strain presenting the most reduced content of these proteins was ATCC 15305 strain (Figure 3, Panel A). In order to validate this finding, we performed urease enzymatic assay in the

protein extracts from the three analyzed strains. The result depicted in the Figure 3, panel B corroborates the findings of the proteomic analysis. The strain 7108 presented a very increased urease activity when compared to the other strains. Comparison among 9325 and 15305 strains shows that the first one presents higher level of urease activity. Since urease is a secreted protein, we decide to investigate the urease activity outside the cell. We performed urease activity assay of secreted urease in agar plates. The *S. saprophyticus* cells were serially diluted (10^2 to 10^5 cells) and inoculated in urease agar plates. The result is shown in the Figure 3, panel C. The results show that the ATCC15305 present low urease activity inner (0.1 AU/assay) and outside the cell (detected only in 10^4 and 10^5 cell dilutions), reflecting probably a lower production of the enzyme. Although 7108 strain presents a high urease activity inner the cell, the urease activity outside the cell is low (detected only in 10^4 and 10^5 cell dilutions), suggesting this enzyme is accumulating in the cytoplasm and is not efficiently secreted by this strain. In counterpart, the 9325 strain does not accumulate urease inner the cell and the secretion system is more efficient, since the urease activity inside the cell is low and the urease activity outside the cell is very high, being detected in all the cell dilutions tested.

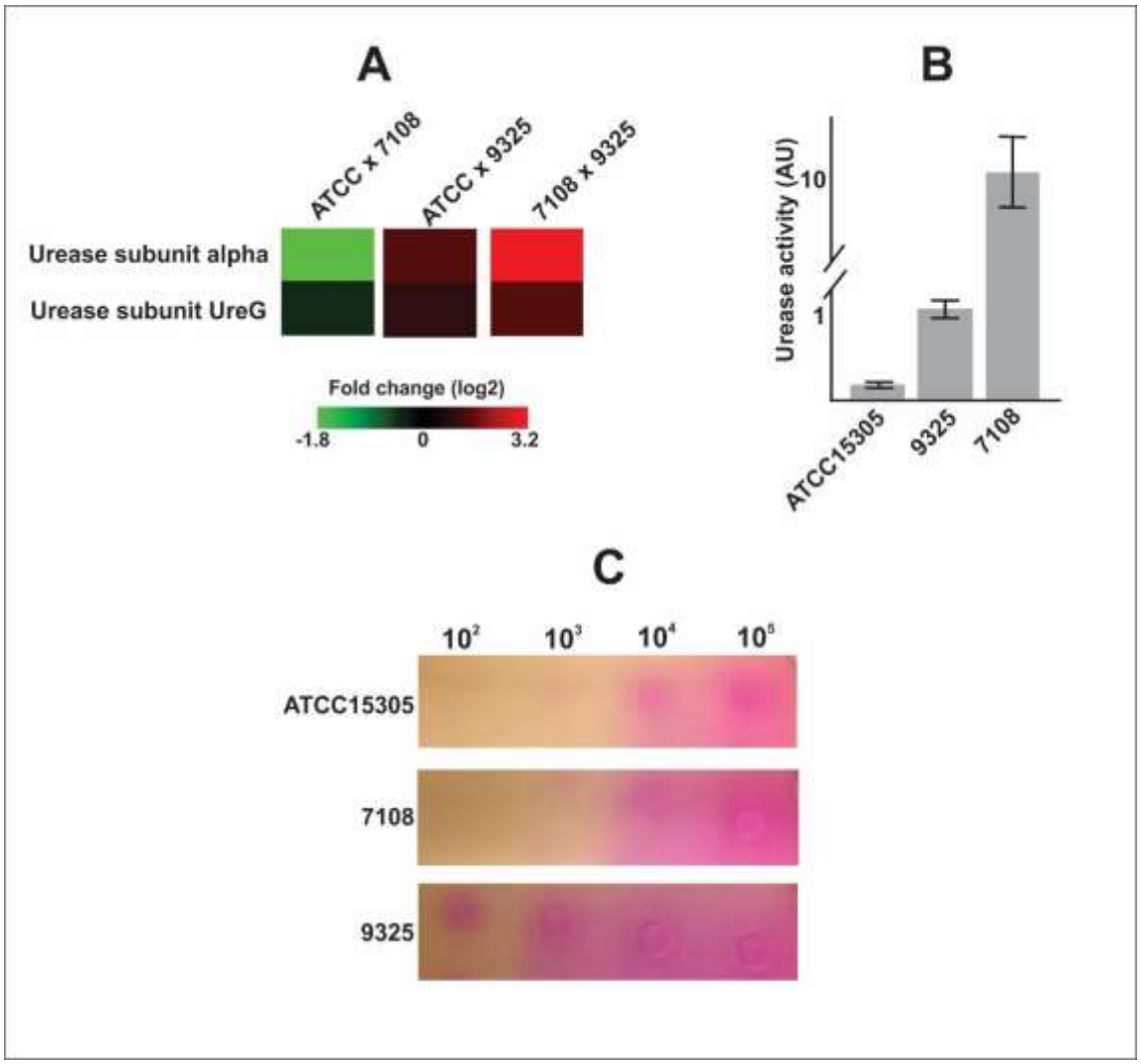


Figure 3 – Urease enzymatic assays with protein extract from and secreted by *S. saprophyticus* cells. (A) Heat map of urease subunits expression. Heat map showing fold change (log2) comparing urease subunits expression among *S. saprophyticus* strains. The heat map was generated using proteomic data. **B: Urease activity**. The enzymatic assay is shown in arbitrary units (AU) performed with protein extracts of *S. saprophyticus* strains. The experiment was performed using biological replicate and with technical triplicates. **B: Evaluation of secreted urease activity.** The *S. saprophyticus* cells were serially diluted from 10^5 to 10^2 cells and inoculated inurease agar plates. The urease activity is detected by change of color of the medium, from yellow to purple. The experiment was performed with biological triplicates and representative images are shown.

The *S. saprophyticus* strains used in this work were previously analyzed and also differs in the exoproteome content [14]. In order to compare if the phenotypic, exoproteomic and proteomic differences described for these strains could reflect in the ability to survive during interaction with host cells, we performed interaction assays of *S. saprophyticus* cells with macrophages. The result is shown in the Figure 4, panel A. It is remarkable that 9325 strain presents the higher ability to survive during interaction with macrophages, when compared to ATCC 15305 and 7108 strains. The strain presenting the lowest survival rate to macrophage interaction was 7108. ATCC 15305 survival rate is slight reduced in comparison with 9325 strain. We also evaluated the ability of the *S. saprophyticus* strains to form biofilm in polystyrene plates. The result is shown in the Figure 4, panel B. Surprisingly, it was possible to detect that non-capsular strains 7108 possess the higher ability to form biofilm when cultured in BHI medium containing 1% of glucose.

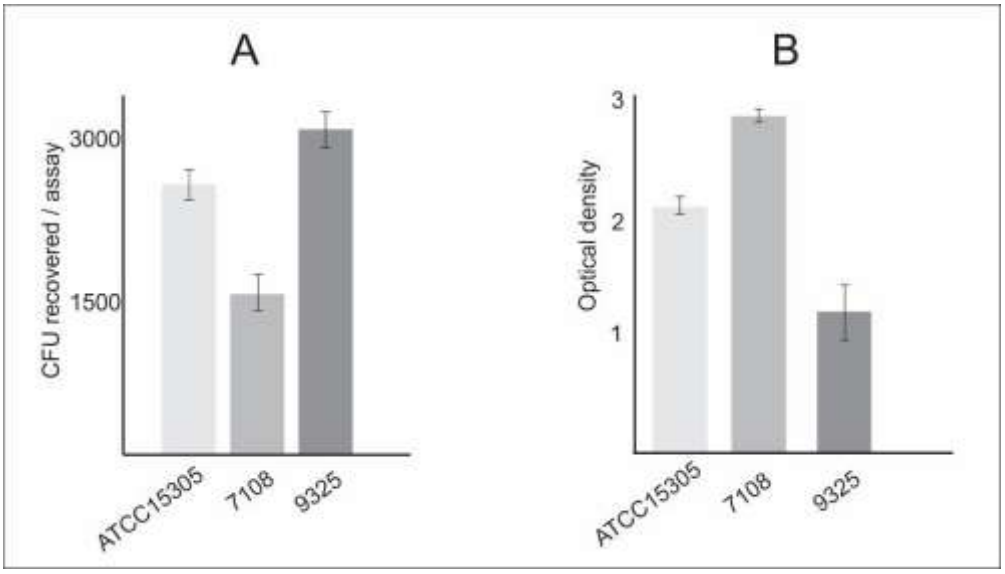


Figure 4 – *S. saprophyticus* interaction assay with macrophages and evaluation of biofilm formation. A: The *S. saprophyticus* cells were incubated with macrophages and, after interaction assay, the macrophages were lysed and CFU

recovered and plated in BHI medium. The experiments were performed in biological triplicate and standard error of the mean was calculated. **B:** The biofilm assay was performed in polystyrene plate, cells were fixed and stained with crystal violet. Optical density was measured at 570 nm wavelength. The experiments were performed in biological triplicate and standard error of the mean was calculated.

In order to summarize phenotypic, proteomic and exoproteomic differences related to pathogenesis and virulence among the strains we schematized the phenotypic and proteins identified in the strains. The scheme is shown in the Figure 5. The schematic figure was generated based on previous work of phenotypic characterization of capsule[13], our previous studies of comparative exoproteome analysis [26] and in the results of this work. It is remarkable that 7108 strain possess a reduced repertoire of proteins related to virulence compared to ATCC 15305 and 9325 strains. The secretion of uroadherence factor (UroA) and proteases was described in exoproteome analysis [14]. The ATCC 15305 strain machinery related to virulence and pathogenesis detected in this work includes highly production of proteins related to purine synthesis, tricarboxylic acid cycle and thioredoxins, besides the previous description of secretion of thioredoxins, UroA and the antigenic protein SsaA [14]. The 9325 machinery involved in virulence seems to be wider when comparing to ATCC 15305 and 7108 *S. saprophyticus* strains. This strain possess secretion of several proteins related to virulence, such as highly amount of urease, SsaA, several antigenic proteins and thioredoxins [14] and we detected the production of highly amounts of thioredoxins, chaperone proteins and proteins related to glycolysis in comparison with the other strains. Some proteins important during infection, such as ferrochelatases and siderophore transporters were detected in this work. Previous analysis from our group demonstrate the relevance of iron metabolism to

increase *S. saprophyticus* survival rate during macrophage infection, reinforcing these proteins are important during infection process [25].

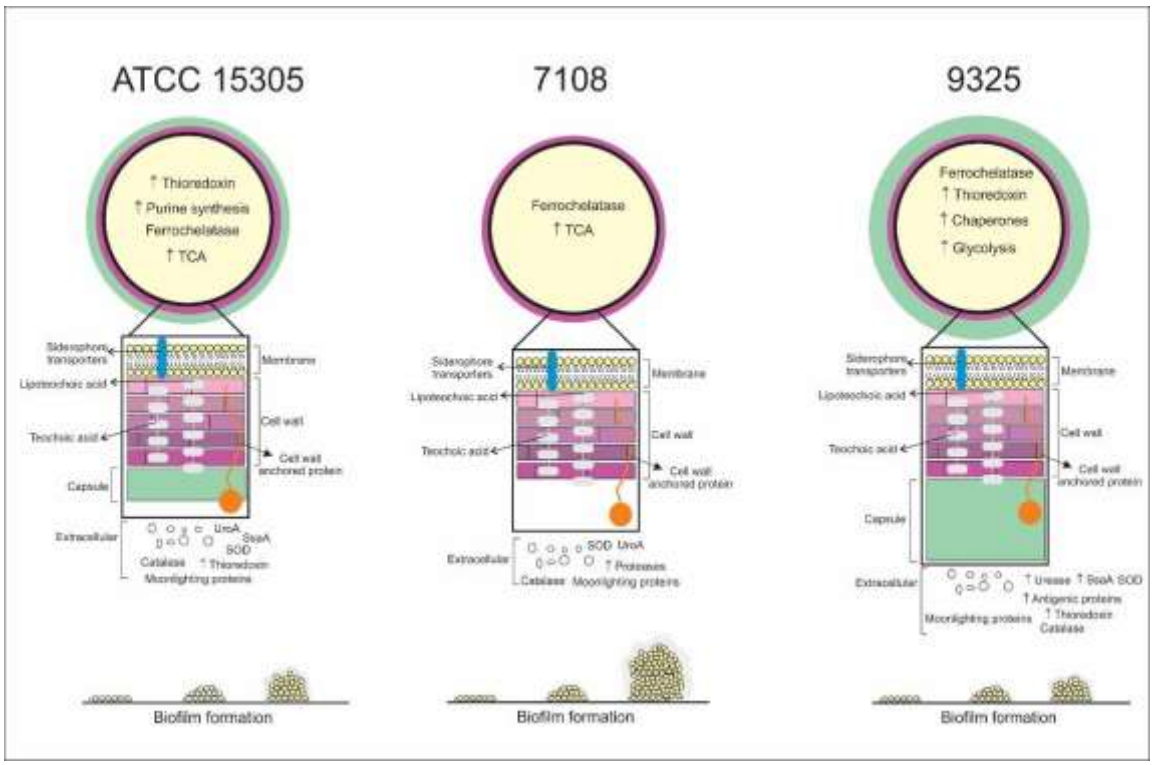


Figure 5 – Schematic representation of phenotypic, exoproteomic and proteomic differences among the *S. saprophyticus* strains analyzed. Proteins and phenotypic differences (detected in previous works and in this work) related to virulence, pathogenesis and persistence are shown. TCA: tricarboxylic acid cycle; UroA uroadherence factor A; SsaA: staphylococcal antigenic protein A; SOD: superoxide dismutases.

DISCUSSION

In the last years, proteomic approaches have been used to understand pathogens machinery used to infect host [18, 19]. Specially, staphylococcal species have been studied by these techniques, providing elucidation of processes used during infection and contributing in the discovery of virulence determinants thus enlarging knowledge about pathogenicity mechanisms [20, 21].

The knowledge about the proteomic profile of *S. saprophyticus* is still incipient. Few analysis using proteomic approach were performed in *S. saprophyticus*, but our previous comparative exoproteome analysis among three *S. saprophyticus* strains (7108 non-capsular strain and ATCC 15305 and 9325 capsular strains) shows diversity in the exoproteome contents, suggesting these strains can possess different machineries of secreted proteins to promote infection [14]. In this work we are describing the proteome global content of these three *S. saprophyticus* strains. The cell growth curves of these strains were evaluated in rich medium and they presented similar growth rate in this condition (Supplementary Figure 1). The proteomic approach was used to obtain the proteome profile of the strains and the proteins identified were functionally categorized (Figure 1). The statistical analysis performed included results obtained from biological triplicates and experimental replicates in order to enhance the resolution of the results obtained. From the total of 276 identified proteins, 170 were not regulated among the strains. It includes proteins involved in protein synthesis, transcription and DNA metabolism (Supplementary Table 1) and reinforce that *S. saprophyticus* cells from the three strains were growth until the same metabolic moment of the cell growth curve before proteomic analysis.

Among the regulated proteins detected in each *S. saprophyticus* strain, we highlighted proteins related to virulence and pathogenesis (Table 1). We detected that 9325 and ATCC 15305 strains possess higher amounts of thioredoxins and reductases in comparison with 7108 strain and this data was confirmed by thiol reduction dosage (Figure 2). Thioredoxin system (Trx) – formed by thioredoxin reductase, thioredoxin and NADPH - can scavenge ROS [22]. In *S. aureus*, the glutathione system is deficient and Trx system is important for bacterial survival under oxidative stress conditions [23]. In this sense, strains possessing higher content of thioredoxins could be related to higher

adaptation and survival under oxidative stress condition. Other difference detected among the *S. saprophyticus* strains is the content of urease. We detected the 7108 strain possess higher content of urease in the proteomic analysis in comparison with ATCC 15305 and 9325 strains. We checked if the higher urease content could be due the accumulation of urease inner the cell or could be due the higher production of this protein in the 7108 strain. The results show that this strain does not secretes urease efficiently and urease remains inner the cell (Figure 3). Urease secretion is considered a secreted virulence factor in *S. ssaprophyticus* contributing to urophatogenicity in rats [3] and facilitating *S. saprophyticus* growth in artificial urine medium [4]. This finding suggests strains 9325 and ATCC 15305 could present higher ability to survive during host interaction since they can export urease with more efficiency when compared to 7108 strain.

It was possible to note that phenotypic and proteomic differences described among these strains reflect in the survival ability during interaction with host cells (Figure 4, panel A) and in the ability to form biofilm (Figure 4, panel B). Our findings show the *S. saprophyticus* 9325 strains possess higher survival rate after interaction with macrophages, followed by ATCC 15305 strain. The lowest survival rate was detected in the 7108 strain. In counterpart, the 7108 presented the highest ability to form biofilm when cultured in BHI medium enriched with glucose. This data corroborates previous results using *Streptococcus pneumoniae* as model. Experiments of biofilm formation using defected mutants describes suggesting that the capsule is antagonistic to biofilm formation [27].

The differences among these strains related to virulence and pathogenicity described in this work and in previous work [13, 14] summarized in the Figure 5 show that 9325 and ATCC 15305 possess larger amounts of proteins that could enhance pathogenicity and virulence. The lowest machinery of proteins that could enhance

virulence is described in the 7108 strain. In counterpart, the ability to form biofilm is increases in the 7108 strain. Although the lower production of virulence factor detected in this and in previous work, the ability to form biofilm can enhance the survival rate in the host since biofilm formation is associated with reduction of *Staphylococcal* cells access to the defense system of the host and is also associated with impairing of antibiotic action [11]. Analysis of clinical *S. saprophyticus* strains described the biofilm formation ability is presented by around 70% of the strains, suggesting it could be important during infection since increases resistance to several antibiotic [12]. The increasing in the persistence rate caused by biofilm is also important since enhance conjugation in staphylococcal species [11]. Our results and analysis show that clinical *S. saprophyticus* strains can possess different protein machinery and phenotypic characteristic that can confer ability to invade and persist in the human host. These results can be important since protein targets for drug development must take into account machinery conserved among clinical strains to be efficient. Thus, it is not enough that the strain possesses the genes associated with virulence and persistence, but the gene must be expressed and used at the time of infection in the host to be a good target. This is the first description of proteomic flexibility among *S. saprophyticus* strains reflecting in virulence, pathogenicity and persistence.

DISCLOSURES

This work was supported by Fundação de Amparo à Pesquisa do Estado de Goiás (FAPEG, Pronex) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) This work is part of Instituto Nacional de Ciência e Tecnologia da Interação Parasito-Hospedeiro (INCT-IPH). KCSS was supported by scholarship from Universidade Federal de Goiás and LOHSS was supported by scholarship from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

1. **Raz R, Colodner R, Kunin CM. Who are you--*Staphylococcus saprophyticus*? *Clin Infect Dis* 40(6), 896-898 (2005).**
2. **Kloos WE, Bannerman TL. Update on clinical significance of coagulase-negative staphylococci. *Clinical microbiology reviews* 7(1), 117-140 (1994).**
3. **Gatermann S, John J, Marre R. *Staphylococcus saprophyticus* urease: characterization and contribution to uropathogenicity in unobstructed urinary tract infection of rats. *Infection and immunity* 57(1), 110-116 (1989).**
4. **Loes AN, Ruyle L, Arvizu M, Gresko KE, Wilson AL, Deutch CE. Inhibition of urease activity in the urinary tract pathogen *Staphylococcus saprophyticus. Letters in applied microbiology* 58(1), 31-41 (2014).**
5. **Gatermann S, Meyer HG, Marre R, Wanner G. Identification and characterization of surface proteins from *Staphylococcus saprophyticus. Zentralbl Bakteriol* 278(2-3), 258-274 (1993).**
6. **Korte-Berwanger M, Sakinc T, Kline K, Nielsen HV, Hultgren S, Gatermann SG. Significance of the D-serine-deaminase and D-serine metabolism of *Staphylococcus saprophyticus* for virulence. *Infection and immunity* 81(12), 4525-4533 (2013).**
7. **Argemi X, Hansmann Y, Prola K, Prevost G. Coagulase-Negative Staphylococci Pathogenomics. *Int J Mol Sci* 20(5), (2019).**
8. **Lee JH, Heo S, Jeong M, Jeong DW. Transfer of a mobile *Staphylococcus saprophyticus* plasmid isolated from fermented seafood that confers tetracycline resistance. *PloS one* 14(2), e0213289 (2019).**
9. **Sousa VS, Da-Silva APS, Sorenson L *et al*. *Staphylococcus saprophyticus* Recovered from Humans, Food, and Recreational Waters in Rio de Janeiro, Brazil. *Int J Microbiol* 2017 4287547 (2017).**
10. **Li H, Xu L, Wang J *et al*. Conversion of *Staphylococcus epidermidis* strains from commensal to invasive by expression of the ica locus encoding production of biofilm exopolysaccharide. *Infection and immunity* 73(5), 3188-3191 (2005).**
11. **Aguila-Arcos S, Alvarez-Rodriguez I, Garaiyurrebaso O, Garbisu C, Grohmann E, Alkorta I. Biofilm-Forming Clinical *Staphylococcus* Isolates Harbor Horizontal Transfer and Antibiotic Resistance Genes. *Front Microbiol* 8 2018 (2017).**
12. **Martins KB, Ferreira AM, Pereira VC, Pinheiro L, De Oliveira A, Da Cunha M. In vitro Effects of Antimicrobial Agents on Planktonic and Biofilm Forms of *Staphylococcus saprophyticus* Isolated From Patients With Urinary Tract Infections. *Front Microbiol* 10 40 (2019).**
13. **Kleine B, Gatermann S, Sakinc T. Genotypic and phenotypic variation among *Staphylococcus saprophyticus* from human and animal isolates. *BMC Res Notes* 3 163 (2010).**
14. **Oliveira AS, Rosa IIR, Novaes E *et al*. The exoproteome profiles of three *Staphylococcus saprophyticus* strains reveal diversity in protein secretion contents. *Microbiol Res* 216 85-96 (2018).**
15. **Park S, Kelley KA, Vinogradov E *et al*. Characterization of the structure and biological functions of a capsular polysaccharide produced by *Staphylococcus saprophyticus. Journal of bacteriology* 192(18), 4618-4626 (2010).**
16. **Ritchie ME, Phipson B, Wu D *et al*. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research* 43(7), e47 (2015).**

17. **Phipson B, Lee S, Majewski IJ, Alexander WS, Smyth GK. Robust Hyperparameter Estimation Protects against Hypervariable Genes and Improves Power to Detect Differential Expression. *Ann Appl Stat* 10(2), 946-963 (2016).**

18. **Martinez-Martin N. Technologies for Proteome-Wide Discovery of Extracellular Host-Pathogen Interactions. *J Immunol Res* 2017 2197615 (2017).**

19. **Jean Beltran PM, Federspiel JD, Sheng X, Cristea IM. Proteomics and integrative omic approaches for understanding host-pathogen interactions and infectious diseases. *Mol Syst Biol* 13(3), 922 (2017).**

20. **Hecker M, Mader U, Volker U. From the genome sequence via the proteome to cell physiology - Pathoproteomics and pathophysiology of *Staphylococcus aureus. Int J Med Microbiol* 308(6), 545-557 (2018).**

21. **Bonar E, Wojcik I, Wladyka B. Proteomics in studies of *Staphylococcus aureus* virulence. *Acta Biochim Pol* 62(3), 367-381 (2015).**

22. **Arner ES, Holmgren A. Physiological functions of thioredoxin and thioredoxin reductase. *European journal of biochemistry / FEBS* 267(20), 6102-6109 (2000).**

23. **Uziel O, Borovok I, Schreiber R, Cohen G, Aharonowitz Y. Transcriptional regulation of the *Staphylococcus aureus* thioredoxin and thioredoxin reductase genes in response to oxygen and disulfide stress. *Journal of bacteriology* 186(2), 326-334 (2004).**

23. **Murad, A.M., Rech, E.L., 2012. NanoUPLC-MSE proteomic data assessment of soybean seeds using the Uniprot database. BMC Biotechnol. 12, 82**

24. **A. Soriano, G.J. Colpas, R.P. Hausinger, UreE stimulation of GTP-dependent urease activation in the UreD-UreF-UreG-urease apoprotein complex, Biochemistry. 39 (2000) 12435–12440. doi:10.1021/bi001296o.**

25. **B.S. Vieira de Souza, K.C. Sousa Silva, A.F. Alves Parente, C.L. Borges, J.D. Paccez, M. Pereira, C. Maria de Almeida Soares, M. Giambiagi-deMarval, M.G. Silva Bailão, J.A. Parente-Rocha, The influence of pH on S. saprophyticus iron metabolism and the production of siderophores, Microbes Infect. (2019). doi:10.1016/j.micinf.2019.04.008.**

26. **A.S. de Oliveira, I.I.R. Rosa, E. Novaes, L.S. de Oliveira, L.C. Baeza, C.L. Borges, L. Marlinghaus, C.M. de A. Soares, M. Giambiagi-deMarval, J.A. Parente-Rocha, The exoproteome profiles of three Staphylococcus saprophyticus strains reveal diversity in protein secretion contents, Microbiol. Res. 216 (2018) 85–96. doi:10.1016/j.micres.2018.08.008.**

27. **E.J. Muñoz-Elías, J. Marcano, A. Camilli, Isolation of Streptococcus pneumoniae biofilm mutants and their characterization during nasopharyngeal colonization, Infect. Immun. 76 (2008) 5049–5061. doi:10.1128/IAI.00425-08.**

MANUSCRITO II

The interacting purine *de novo* pathway with carbohydrate biosynthesis proteins modulates *Staphylococcus saprophyticus* biofilm formation

Karla Christina Sousa Silva[1#], Kleber Santiago Freitas e Silva[1#], Raisa Melo Lima[1], Clayton Luiz Borges[1], Marcia Giambiagi-deMarval[2], Célia Maria de Almeida Soares[1], Maristela Pereira[1], Juliana Alves Parente-Rocha[1*]

[1]*Laboratório de Biologia Molecular, Instituto de Ciências Biológicas, Universidade Federal de Goiás, Goiânia, Goiás, Brazil.*

[2]*Laboratório de Microbiologia Molecular, Instituto de Microbiologia Prof. Paulo de Góes, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Rio de Janeiro, Brazil.*

 # These authors equally contributed in this work.

KCSS: karlabio@live.com

KSFS: smallbinho@hotmail.com

RML: raisamelolima@hotmail.com

MGM: marciagm@micro.ufrj.br

CMAS: cmasoares@gmail.com

MP: maristelaufg@gmail.com

JAPR: juparente@ufg.br

* Corresponding author: Juliana Alves Parente-Rocha. Universidade Federal de Goiás. Avenida Esperança, s/n ICB2, Laboratório de Biologia Molecular. 74690-900. Goiânia, GO, Brazil. juparente@ufg.br

ABSTRACT

Staphylococcus saprophyticus is one of the leading causes of uncomplicated urinary tract infections (UTI) worldwide. It mainly affects young sexually active women and it has also been associated with biofilm formation on medical devices. Proteomic analyses have shown a high expression of *de novo* purine biosynthetic pathway in ATCC 15305 strain compared to other two clinical strains. Considering the importance of purine synthesis to virulence in several pathogenic organisms, we evaluated the impact of the suppression of *de novo* purine biosynthetic pathway upon biofilm production in ATCC 15305 strain. Subsequently, we evaluated how natural selection acted upon specific genes of *de novo* purine synthesis pathway. Our results show that once the pathway is overexpressed the bacteria tends to produce more biofilms. The *in silico* molecular interactions suggested that key gene product of *purH* interacts with enzymes responsible for converting cell wall intermediates into substrate for glycolysis, suggesting PurH interaction negatively modulates this interconversion. Therefore, we hypothesize that *purH* may be responding to pressures in order to improve survivability of the pathogen inside hosts. Given the fact that *purH* is absent in humans, it may be an interesting target for rational drug design.

Key words: purine biosynthesis; interactome; biofilm.

INTRODUCTION

Staphylococcus saprophyticus is a gram positive, coagulase-negative and novobiocin resistant bacteria which is known to be one of the leading causes of uncomplicated urinary tract infections (UTI) worldwide [1]. *S. saprophyticus* is more common among sexually active females and it has special urotropic and ecologic features that are distinctly different from other staphylococci and from *Escherichia coli* [2].

One of the most important virulence features of *S. saprophyticus* is the ability to form biofilms. Biofilms are formed by sessile cells attached to a surface which are able to encase themselves in a hydrated matrix of polysaccharide and protein, forming a slimy layer [3] that present considerably resistance to antibiotics [4]. The cells adhered to biofilms display different growth rate, gene expression and protein synthesis compared to their planktonic (floating) counterparts [5].

Due to the advances in medical treatments, the use of prosthetic devices has increased significantly and it has reflected upon the frequency of infections caused by bacterial colonization of peacemakers, breast implants, defibrillators, heart valves, and aortic grafts [6].

Staphylococci contains the *ica* operon which is responsible for synthesizing one of the main biofilm components, the polysaccharide intercellular adhesin (PIA) or polymeric *N*-acetyl-glucosamine (PNAG). Several intricate mechanisms act upon the biofilm modulation. In *Staphylococcus epidermidis* the adhesion is negatively regulated by *agr* quorum sensing system. The transcription of *ica*ADBC is modulated by genes products of *purR* (negative controller of purine nucleobase metabolic process), *sarA* (regulation of transcription) and the *luxS* quorum-sensing system; IcaR works as a repressor of *ica*ADBC transcription [7]. By contrast, *nagABCD* operon encode genes

which use GlcNAc6P as a substrate for conversion into GlcN6P and the further step converts GlcN6P into Fructose-6P [8]. Previous studies described that purine biosynthesis is an important pathway for virulence of pathogenic bacteria, such as *Bacillus anthracis* [9], the uropathogenic *Escherichia coli* [10] and other staphylococci species [11].

Nucleotides are synthesized through two different pathways: salvage and *de novo*. The salvage pathway recycles the nucleosides and free bases derived from nucleic acid degradation. The *de novo* pathway synthesizes purines and pyrimidines using available molecular precursors not related to nucleotides. The genes from this pathway are organized in an operon in Staphylococci genomes which contains eleven genes (*purD, purH, purN, purM,purF, purQ, purL, purS, purC, purK, purE*). The disruption of the *de novo* purine synthesis pathway has an impact on biofilm formation [9], growth [10], and virulence [12].

Comparative proteome analysis of *S. saprophyticus* shows that ATCC15305 strain expresses *pur* operon genes in a higher rate than another two clinical isolates (7108 and 9325) [13]. Therefore, we aimed to evaluate which effects the inactivation of *pur* operon would exert upon the biofilm modulation in these strains; predict which gene products interact with key proteins from *de novo* synthesis to form biofilms and how these genes interactions have involved throughout the evolution of the species.

MATERIAL AND METHODS

Prediction of purine *de novo* biosynthesis operon

The *in silico* analysis for prediction of the operon related to purine *de novo* biosynthesis was performed by using the Operon prediction online tool (http://operondb.cbcb.umd.edu/cgi-bin/operondb/taxon_list.cgi). Homology analysis was performed using the NCBI blast tool (https://blast.ncbi.nlm.nih.gov/Blast.cgi).

Prediction of interaction network

Proteins from the *S. saprophyticus* purine *de novo* biosynthesis operon were submitted to STRING server (https://string-db.org/) for evaluation of interaction network. The search was performed based of experimental evidence with a minimal score 0.15.

Determination of the evolutionary history of purB, purE and pur H of *S. saprophyticus*

We retrieved the nucleotide sequence of 51 fungal species from The National Center for Biotechnology Information (NCBI). The inclusion criteria for the analysis was based on sequences within at least 30% identity to the *S. saprophyticus* PurB, PurE and PurH amino acid sequence. The evolutionary history was inferred using the Neighbor-Joining method [14]. The bootstrap consensus tree was inferred from 500 replicates and represented the evolutionary history of the taxa analyzed [15]. The evolutionary distances were calculated through the maximum composite likelihood method [16] and were set in the units of the number of base substitutions per site. Codon positions included were 1st+2nd+3rd+Noncoding. All ambiguous positions were removed for each sequence pair in a pairwise deletion manner. We considered a number of 400 positions in the final dataset. Evolutionary analyses were conducted using MEGA X software [17].

Biofilm crystal violet staining assay

For biofilm assay we performed a modified protocol of crystal violet (CV) staining test previously described [18]. The *S. saprophyticus* cells from strain ATCC 15305 were cultured in BHI (Brain Heart Infusion, Sigma Aldrich, St. Louis, MO, USA) broth for 24 h at 37°C, then washed three times with phosphate saline buffer (PBS), transferred to M9 medium without purines or M9 medium [10] supplemented with 40 µM Adenine and 40µM Guanine and diluted to 1:100. These dilutions were placed in a final volume of 200

µl into 96-well polystyrene microtiter plates (Cralplast, São Paulo, SP, Brazil) and incubated for another 24 h at 37°C. After removal of bacterial suspension, the wells were washed twice with PBS and the biofilm fixed for 1 h at 60°C. The 0.1 % CV solution was added to each well in a final volume of 200 µl, after 15 minutes the solution was rinsed with water and dried 30 minutes at 60°C. Surface-bound CV was extracted by the addition of 200 µl 95% ethanol and incubated for 30 minutes and optical density measured at 570 nm wavelength using spectrophotometry SpectraMax Paradigm (Molecular Devices, Lagerhausstrasse, Austria).

Natural selection analysis for *S. saprophyticus* PurH protein

We performed a SLAC (single likelihood ancestor counting) prediction with a maximum likelihood ancestral state build and a minimum path substitution in order to estimate dS (synonymous rate) and dN (non-synonymous rate) at an amino acid residue level. A binomial test was used in order to compare the dS/dN ratio. Statistical analyzes were used to obtain information over all nucleotide sequence branches, signaling pervasive diversification or conservation in order to define either a positive or a negative selection at a specific residue site. The natural selection analysis and the dN/dS ratio test were performed by Datamonkey Adaptive Evolution Server (https://www.datamonkey.org/). We chose Nucleotide GTR and Global MG94xREV to compute predictive scores for the natural selection analysis, based on substitution models. The former is the most general and consider independent variables, with parameters for nucleotides based on a frequency vector; the latter is based on a codon model allowing for the generation of a full mutation rate matrix [19].

Conformational structures prediction of PurH, NagC and NagD proteins from *S. saprophyticus*

The conformational structure of the proteins purH, NagC and NagD were modeled by the I-TASSER (Iterative Threading Assembly Refinement) server [20]. The modeling is based on homology from three-dimensional protein structures (2bmb and 5z79) experimentally resolved and deposited into the protein databank (PDB - https://www.rcsb.org/) . The final structure is assembled by Monte Carlo simulations. The methodology comprises six steps. The secondary structure is predicted by PSSpred (Protein Secondary Structure Prediction) and the templates are selected by LOMETS (Local Meta-Threading-Server) [21]. Ranked homologous fragments is scored and assembled according to Monte Carlo simulations [22]; conformation and energy scores are used to cluster the ranked fragments in order to propose a model similar to the proteins native structures [23]; molecular dynamics conformational enhancement and the prediction of biological function by the COACH server (https://zhanglab.ccmb.med.umich.edu/COACH/) [24] completes the modeling approach.

The domains of the proteins were identified by KBDOCK [25] and InterPro [26]. Protein-protein docking were performed by ClusPro [27], through clusterization and minimization of the predicted models. The protein-protein interaction (PPI) and the interface of interaction between the target proteins were build based on balanced-favored coefficients of thermodynamics. The predicted PPIs were optimized according to energy scores based on those coefficients.

The visualization software PyMol (https://pymol.org) was used in order to analyze the PPI results, the interface of interaction between purH, NagC and NagD and the predicted hot spots. Amino acid residues that significantly contribute to the free-energy of binding and stability of PPI within the interface of interaction were recognized by KFC2 [28]. The identification of hot spot residues was based on structural and chemical

analysis of the environment around amino acid residues within the interface of interaction. In addition, hot spots experimentally determined by alanine scanning mutagenesis were taken into account for the prediction of hot spot present in the proteins under investigation. The hot spot prediction scores were based on conformation and biochemical properties.

RESULTS

Description of *S. saprophyticus* operon for purine biosynthesis

As stated above purine biosynthesis is an important pathway for virulence of pathogenic bacteria [8, 9, 10]. Is important to note that there is no data about the organization of the *S. saprophyticus* operon. In this sense, we performed *in silico* analysis to elucidate the *S. saprophyticus* operon, using genomic data from ATCC 15305 strain. The Figure 1, panel A shows the purine biosynthesis pathway, conserved among all kingdoms, as previously described [29]. In the Figure 1, panel B it is shows the *S. saprophyticus* operon organization for purine biosynthesis. The description of operon genome localization, gene names and protein names is shown in the Table 1. Of special note, we did not detect the *purB* gene - required in the purine biosynthesis - in the *S. saprophyticus* operon. All the other genes required for inosine monophosphate (IMP) production are localized in the operon.

Table 1 – Description of the genes composing purine *de novo* biosynthesis operon from *S. saprophyticus*.

Locus	Localization in the genome (bp)	Gene name	Protein name

SSP1716	1783762-1785006	*purD*	GAR synthetase / PurD
SSP1717	1785031-1786509	*purH*	AICAR transformylase / PurH
SSP1718	1786525-1787091	*purN*	GAR transformylase / PurN
SSP1719	1787094-1788122	*purM*	AIR synthetase / PurM
SSP1720	1788115-1789602	*purF*	Amidophosphoribosyltransferase / PurF
SSP1721			
SSP1722	1789581-1792694	*purQLS*	FGAM synthetase II / PurQLS
SSP1723			
SSP1724	1792697-1793398	*purC*	SAICAR synthase / PurC
SSP1725	1793403-1794530	*purK*	NCAIR synthetase / PurK
SSP1726	1794517-1794999	*purE*	AIR carboxylase / PurE

Protein AICAR transformylase (PurH) interacts with proteins related to glucosamine metabolism

In order to detect proteins interacting with the proteins produced by the *S. saprophyticus* purine biosynthesis operon, we performed interaction network analysis, as described above. The interaction networks of proteins PurD, PurN, PurM, PurF, PurQ, PurL, PurS, PurC and PurE are shown in the Figure 2. It was not possible to detect interactions occurring with PurK protein using the parameters described above in the STRING database. The list of proteins interacting with proteins of the purine biosynthesis operon is shown in the Supplementary Table 1. Several proteins synthesized by the *pur* operon interact with each other, such as PurC, PurD and PurL. It can optimize the purine synthesis since avoid deviation of metabolites to other pathways [30].

Of special note, the Figure 3 shows the interaction network of the PurH protein. The PurH protein catalyzes the penultimate and the last steps of the purine biosynthesis, forming IMP [29]. It was possible to detect that PurH interacts with 10 proteins, as described in the Table 2. We highlighted the interaction of PurH with NagB and NagD and 2 glucosamine 6 phosphate isomerase (GAPI) proteins. The operon composed *nag*ABCD is responsible for production of NagABCD proteins that converts N acetyl glucosamine 6-phosphate to fructose 6 phosphate [31] to be use by glycolysis. NagB is a glucosamine-6-phosphate deaminase that catalyzes the formation of fructose 6 phosphate from glucosamine 6 phosphate (GAPI catalyzes the same reaction) while NagD function is still unknown [32]. PurH protein also interacts with PurC – from purine biosynthesis operon - and GuaB, that participates of guanine biosynthesis that occurs after IMP production.

Table 2 – Description of proteins that interacts with PurH protein in *S. saprophyticus*.

Gene name	Protein description	Biological process
nagB	glucosamine-6-phosphate deaminase	Glucosamine metabolism
nagD	repressor of the nagBACD operon	Glucosamine metabolism
SSP0235	Glucosamine-6-phosphate isomerase	Glucosamine metabolism
SSP0941	Glucosamine-6-phosphate isomerase	Glucosamine metabolism
deoC	Deoxyribose-phosphate aldolase	Catabolism of deoxyribonucleotides
guaB	Inosine monophosphatedehydrogenase	GMP biosynthesis
lysS	Lysine-tRNA ligase	Protein synthesis
purC	SAICAR synthase	Purine biosynthesis

SSP1416	Tranketolase	Pentose phosphate pathway
SSP1552	Putative kinase	--

Effect of nucleotide metabolism on the biofilm formation in *S. saprophyticus*

Our *in silico* analysis shows PurH interacts with proteins related to glucosamine utilization for energetic purpose. If the interaction positively regulates this pathway, it may reduce glucosamine availability for synthesis of structural polysaccharides. In counterpart, if PurH regulation is negative for utilization of glucosamine in the energy metabolism, could reflect the increased availability of glucosamine for formation of capsule, cell wall and biofilm polysaccharides. In order to evaluate the effect of PurH interaction in the biofilm formation, we performed biofilm assay with *S. saprophyticus* cells after culturing bacterial cells with and without purines. The result is shown in the Figure 4, showing the overexpression of purine synthesis - that occurs when bacterial cells are cultured without purine – positively regulates the biofilm formation at a rate of 90%, approximately. This result, associated with *in silico* analyzes suggest that PurH negatively modulates the glucosamine utilization by energy metabolism, increasing the glucosamine availability for biofilm formation.

Phylogenetic analysis shows conservation of motifs in *S. saprophyticus purH* gene

In order to analyze the diversity of the phylogenetic clades, the well-documented structural features of the *purH* gene, were checked for their presence in the sequence alignment. We found MGS (methylglyoxal synthetase) and AICARFT (5-aminoimidazole-4-carboxamide ribonucleotide formyltransferase) domains within PurH protein sequence [33]. Both motifs are related to purine biosynthesis [34, 35].

The phylogenetic classification based on *purH* gene is shown in the Supplementary Figure 1. It was constructed with 51 nucleotide sequences encoding to *purH* gene of several species with identity rate of 30–95%. Branches with bootstrap values >50% was considered fairly reliable.

S. saprophyticus PurH is under positive natural selection

Natural selection processes control adaptive changes that all living organisms undergo. We performed natural selection analysis in order to determine if PurH protein sequence is under positive selection, neutral or negative selection based on nucleotide or amino acid residues evaluation [19]. We achieved strong negative values for the log likelihood scores of Nucleotide GTR and Global MG94xREV regarding PurH (Table 3). The negative values confirm that both models had fitted the data consistently and the natural selection analysis is reliable, with *p* value threshold of 0.1.

Table 3 – Predictive score for the natural selection analysis.

	Model	AIC_C*	log L	Parameters**
PurH	Nucleotide GTR	143832.86	-71809.16	107
	Global MG94xREV	132086.73	-66043.36	114

* Best-fitting rate matrix value

** Substitution parameters

At this selected threshold, we found positive selection for PurH (Figure 5) at two sites (336 and 276) and negative selection at 364 sites, suggesting amino acids 336 and 276 could be responding to environmental pressures.

Experimental methodologies along with *in silico* approaches have contributed to increase the knowledge regarding pathogens, development of new diagnosis and therapeutic strategies [36, 37]. The prediction of hot spots residues that contributes to stabilize protein complexes is a promisor approach for understanding pathogens and diseases better [37, 38, 39]. Amino acid residues are somewhat conserved among proteins with similar function and proteins that are found in the same complex.

Figure 6 shows the prediction of the structures of purH (A and B), NagC (C and D) and NagD (E and F). The most representative model of interaction between PurH and NagC and the interface of interaction between the proteins under study are represented in Figure 7. The final model of interaction was based on coefficients of energy, such as electrostatic, hydrophobic and Van der Waals-favored.

The interface of interaction of PurH-NagC and PurH-NagD contains hydrophobic amino acid residues, which lead to an increased hydrophobic effect. This effect in protein interactions is the main causes of hot spot residues clusterization [40]. We found cluster of hot spots for both interactions of PurH. The main differences between these two predicted states are the hot spots residues identified for each situation, even though some of these hot spots repeat within the interface for both interactions. Several studies have shown how electrostatic forces contribute to PPIs [41, 42] including those related to diseases [43].

Clusterization drives hot spot prediction. In fact, we found that our predicted hot spots in the interface of interaction between the proteins are near one to another, forming clusters and contributing for the stability of the complex. Next, we analyze the hot spot residues that most contribute for the stabilization of the complex in each model of interaction. We identified nine main hot spot residues for the interaction between PurH and NagD and only three between PurH and NagC (Table 4). These amino acid residues

establish polar contact with other hot spots present within the cluster and also with neighbors, less important residues for the interaction (Figure 8).

Table 4 – Hotspot residues that significantly contribute to the free-energy of binding through balanced coefficients of energy

Interface	Residue	Score[a]	Score[b]
NagC	GLU 55	0.36	0.04
NagC	ARG 147	1.38	0.29
NagC	ARG 149	0.52	0.06
NagD	LEU 57	1.37	0.01
NagD	TRP 29	1.28	0.29
NagD	ILE 51	1.21	0.04
NagD	GLU 68	1.14	0.22
NagD	ARG 79	0.47	0.07
NagD	LEU 226	0.99	0.02
NagD	GLU 235	0.67	0.50
NagD	LYS 243	1.11	0.33
NagD	PHE 248	0.87	0.12

Score[a] – Score based on conformation

Score[b] – Score based on biochemical properties

The Figure 8, Panel A shows the hot spots prediction for the interaction between PurH and NagC. Interestingly, ARG residues participate in polar interactions in the interface between these two proteins. ARG greatly contributes to binding of toxin proteins and ion channel proteins through electrostatic forces [44] and they can act as electrostatic

adhesive forces among biomolecules [45]. In the present approach, Arg hot spot residues establish polar interactions with neighbor amino acids from the same polypeptide chain and with amino acids from the polypeptide chain of the interacting protein. Thus, it greatly influences the conformation stability of the PurH-NagC complex.

The residue TRP 29, present in the PurH structure, interacts with GLU 235 (Figure 8B) and neighbor residues on of the NagD chain, including ARG (not shown). The Arg side chain is amphipathic and generally found on the surface of several proteins. The hydrophilic part of ARG is able to interact with polar residues present in the interface, aiding the stabilization of the complex [46]. Cluster of hot spots such the ones predicted here, and contribute to the free-energy of the protein complex and has an important role for the biological function of the complex [47].

DISCUSSION

The purine metabolism is well described is several bacterial species associated with virulence, such as *E. coli* and *Staphylococcus aureus* [10, 48]. However, there is no description of how the purine biosynthesis pathway can contribute in the ability of *S. saprophyticus* to survive and persist as pathogen. In this work, we presented the genomic organization of purine biosynthesis operon (Figure 1 and Table 1). In order to identify the interactome of purine biosynthesis operon, we performed the STRING analysis (Figures 2 and 3) and we identified several proteins produced by *pur* operon interacting with each other. The interaction of proteins of the same operon is described and can be part of a "metabolon", as previously described [49]. The complex formed by proteins of the same pathway channels the secondary metabolites from one enzyme to another. In this case, the enzymatic activity is optimized since the deviation of metabolites to other pathways is minimal.

We highlighted the interactions identified with PurH protein (Figure 3, Table 3). Our *in silico* analysis of PurH interaction demonstrates this protein probably interacts with NagB, NagD and 2 GAPI proteins. NagB and GAPI promotes glucosamine conversion to fructose 6 phosphate to be used by energetic metabolism [50]. We hypothesized these interactions could negatively regulates glucosamine utilization for energetic purposes to enhance glucosamine availability for structural polysaccharides production. In order to validate this hypothesis we performed biofilm assay after induction of purine biosynthesis and the result shows the biofilm formation in *S. saprophyticus* (Figure 4). The biofilm formation in staphylococcal species is extremely associated with glucosamine availability since poly N-acetyl (1 e 6) β-glucosamine is a vital constituent of biofilm, together with proteins and DNA [51]. The biofilm formation is dependent of the capsular polysaccharide-adhesin (PSA) that mediates cell adhesion to biomaterials. In *S. epidermidis* the PSA is composed by 15% of glucosamine [52]. In addition, *E. coli* curli extracellular surface fibers – that contribute to biofilm formation – are regulated by glucosamine availability [53].

Evolutionary mechanisms act on a molecular level. Interpreting the evolutionary pressure nucleotide or proteins are submitted to, increase our knowledge on how genes evolved and respond to environmental pressure. Natural selection will either favor or prevent the preservation of genetic variation at a specific locus [54, 55]. Here, we show the conservation of motifs present in the sequences analyzed across different clades, which indicates their critical roles related to structure and function regarding purine biosynthesis performed by *purH*.

A useful way to determine the influence of natural selection at the molecular level is performing the dN/dS ratio (ω) test or even dN-dS. The number of non-synonymous substitutions relative to the number of synonymous substitutions indicates if a certain

group of genes, and consequently a protein, is undergoing positive, neutral or negative selection [56, 57]. Natural selection analyses also highlights environmental response acting at the level of specific amino acid residues, such as SNPs that can increase disease susceptibility or hot spot residues that contribute to the overall free-energy of protein-protein interactions (PPIs) [55, 58]. To our knowledge there has not been a study conducted on the natural selection analysis for the *purH* gene in *S. saprophyticus*.

We hypothesize that the sites 336 and 276 in the PurH sequence are important regions of variation of several bacterial PurH and could significantly contribute to the function of the protein and stabilization of PPIs established by PurH. Those sites correspond to the AICARFT domain, which is directly related to purine biosynthesis and indirectly related to biofilm formation. Interestingly, PurH is regulatory and a key enzyme of the purine biosynthesis pathway in *S. saprophyticus* and could be responding to environmental pressures, including the harsh host environment, in a way that would influence virulence, biofilm formation and survivability of the pathogen inside hosts.

Our results open perspective to assess *in silico* mutation at site 336 and perform dynamical simulation in order to check if this specific site is essential for the conformational structure and function of PurH. This was the first step on elucidation of PurH protein in the modulation of biofilm in *S. saprophyticus*.

DISCLOSURES

This work was supported by Fundação de Amparo à Pesquisa do Estado de Goiás (FAPEG, Pronex) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) This work is part of Instituto Nacional de Ciência e Tecnologia da Interação Parasito-Hospedeiro (INCT-IPH). KCSS was supported by scholarship from Universidade Federal de Goiás. RML and KSFS were supported by scholarship from CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

FIGURE LEGENDS

Figure 1 – Purine biosynthesis pathway and genomic organization of the purine biosynthesis operon from *S. saprophyticus*. A: Purine biosynthetic pathway. The enzyme names are shown in the Table 1. PRPP: 5-phosphoribosylpyrophosphate; PRAM: 5-phospho-D-ribosylamine; GAR: glycinamide ribonucleotide; Formyl-GAR: N-formylglycinamide ribonucleotide; FGAM: N-formylglycinamidine ribonucleotide; AIR: aminoimidazole ribonucleotide; NCAIR: N^5-carboxyaminoimidazole ribonucleotide; CAIR: carboxyaminoimidazole ribonucleotide; SAICAR: N-succinocarboxamide-5-aminoimidazole ribonucleotide; AICAR: aminoimidazole-4-carboxamide ribonucleotide; Formyl-AICAR: 5-formamido-4-imidazolecarboxamide ribonucleotide; IMP: inosine monophosphate. **B:** Genetic map of the purine biosynthesis operon. Open reading frames are indicated by arrows that show the direction of transcription. Size is indicated within the arrows (in base pairs).

Figure 2 – Interaction networks of the proteins produced by purine biosynthesis operon from *S. saprophyticus*. The interaction network of the proteins PurD, PurN, PurM, PurF, PurQ,PurL, PurS, PurC and Pur E is shown. It was not detected interaction occurring with PurK in the STRING database (using experimental evidence as parameter).

Figure 3 - Interaction networks of the PurH protein from *S. saprophyticus*. Interaction of PurH with 10 proteins were identified. Among the identified proteins, 4

proteins are related with glucosamine metabolism (NagB, NagD, SSP0235 and SSP0941) and three of them catalyzes the same reaction (NagB, SSP0235 and SSP0941).

Figure 4 – Evaluation of purine synthesis metabolism on the biofilm formation in *S. saprophyticus.* The ability to form biofilm was evaluated in *S. saprophyticus* after incubation of bacterial cells without (- Pur) and with (+ Pur) purines. The experiment was performed in triplicate and standard error of the mean was obtained.

Figure 5 – Natural selection pressure on *S. saprophyticus* PurH protein. The graphic shows that two sites in the sequence of *S. saprophyticus* PurH is under positive selection (dN-dS>0), with a statistically significant score. Residues at sites 336 and 276 in PurH vary within the nucleotide sequences analyzed. This might be a response to environmental pressure.

Figure 6 – Conformational structures of the proteins under study modeled by I-TASSER. A: PurH cartoon representation; **B:** PurH surface representation; **C:** NagC cartoon representation; **D:** NagC surface representation; **E:** NagD cartoon representation; **F:** NagD surface representation.

Figure 7 - Models of PurH-NagC and PurH-NagD interactions according thermodynamics coefficients. A: Model of interaction between PurH and NagC represented by cartoon. **B:** PurH and NagC interaction in a surface representation. **C:** Model for PurH and NagD interaction regarding represented by cartoon. **D:** PurH and NagD interaction in a surface representation. Salmon: PurH; blue: NagC; orange: NagD; green: interface of interaction between proteins under study.

Figure 8 - Hot spots prediction for the models of PurH interaction with NagC and NagD. All figures presented here show polar interactions for the amino acid residues

classified as hot spots (Table 4). Yellow: hot spot residues; salmon: PurH; blue: NagC; orange: NagD.

REFERENCES

[1] Schito GC, Naber KG, Botto H, Palou J, Mazzei T, Gualco L, et al. The ARESC study: an international survey on the antimicrobial resistance of pathogens involved in uncomplicated urinary tract infections. *Int J Antimicrob Agents.* 2009;34:407-13.

[2] Raz R, Colodner R, Kunin CM. Who are you--*Staphylococcus saprophyticus*? Clin *Infect Dis.* 2005;40:896-8.

[3] Stewart PS, Costerton JW. Antibiotic resistance of bacteria in biofilms. *Lancet.* 2001;358:135-8.

[4] Ceri H, Olson ME, Stremick C, Read RR, Morck D, Buret A. The Calgary Biofilm Device: new technology for rapid determination of antibiotic susceptibilities of bacterial biofilms. *J Clin Microbiol.* 1999;37:1771-6.

[5] Donlan RM, Costerton JW. Biofilms: survival mechanisms of clinically relevant microorganisms. *Clin Microbiol Rev.* 2002;15:167-93.

[6] Darouiche RO. Treatment of infections associated with surgical implants. *N Engl J Med.* 2004;350:1422-9.

[7] Mack D, Rohde H, Harris LG, Davies AP, Horstkotte MA, Knobloch JK. Biofilm formation in medical device-related infection. *Int J Artif Organs.* 2006;29:343-59.

[8] Plumbridge J. An alternative route for recycling of N-acetylglucosamine from peptidoglycan involves the N-acetylglucosamine phosphotransferase system in *Escherichia coli*. *J Bacteriol.* 2009;191:5641-7.

[9] Jenkins A, Cote C, Twenhafel N, Merkel T, Bozue J, Welkos S. Role of purine biosynthesis in *Bacillus anthracis* pathogenesis and virulence. *Infect Immun.* 2011;79:153-66.

[10] Shaffer CL, Zhang EW, Dudley AG, Dixon BR, Guckes KR, Breland EJ, et al. Purine Biosynthesis Metabolically Constrains Intracellular Survival of Uropathogenic *Escherichia coli. Infect Immun.* 2017;85.

[11] Verma P, Kar B, Varshney R, Roy P, Sharma AK. Characterization of AICAR transformylase/IMP cyclohydrolase (ATIC) from *Staphylococcus lugdunensis*. *FEBS J.* 2017;284:4233-61.

[12] Andersen-Civil AIS, Ahmed S, Guerra PR, Andersen TE, Hounmanou YMG, Olsen JE, et al. The impact of inactivation of the purine biosynthesis genes, purN and purT, on growth and virulence in uropathogenic *E. coli*. *Mol Biol Rep.* 2018;45:2707-16.

[13] Silva KCS, Silva LOS, Silva GAA, Borges CL, Novaes E, Paccez JD, Fontes W, Giambiagi-deMarval M, Soares CMA, Parente-Rocha JA. *Staphylococcus saprophyticus* proteomic analyses elucidate differences in the protein repertories among clinical strains related to virulence and persistence. *Pathogens.* 2020;19;9(1):69.

[14] Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular biology and evolution. 1987;4:406-25.

[15] Felsenstein J. Confidence Limits on Phylogenies: An Approach Using the Bootstrap. *Evolution.* 1985;39:783-91.

[16] Tamura K, Nei M, Kumar S. Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc Nat Acad Sci USA.* 2004;101:11030-5.

[17] Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol.* 2018;35:1547-9.

[18] Stepanovic S, Vukovic D, Hola V, Di Bonaventura G, Djukic S, Cirkovic I, et al. Quantification of biofilm in microtiter plates: overview of testing conditions and practical recommendations for assessment of biofilm production by staphylococci. *APMIS.* 2007;115:891-9.

[19] Kosakovsky Pond SL, Frost SD. Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol.* 2005;22:1208-22.

[20] Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER Suite: protein structure and function prediction. *Nat Methods.* 2015;12:7–8.

[21] Wu S, Zhang Y. LOMETS: a local meta-threading-server for protein structure prediction. *Nucleic Acids Res.* 2007;35:3375–3382.

[22] Swendsen null, Wang null. Replica Monte Carlo simulation of spin glasses. *Phys Rev Lett.* 1986;57:2607–2609.

[23] Zhang Y, Skolnick J. SPICKER: a clustering approach to identify near-native protein folds. *J Comput Chem.* 2004;25:865–871.

[24] Yang J, Roy A, Zhang Y. Protein-ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. *Bioinformatics.* 2013;29:2588–2595.

[25] Ghoorah AW, Devignes MD, Smaïl-Tabbone m, Ritchie DW. Classification and Exploration of 3D Protein Domain Interactions Using Kbdock. *Methods Mol Biol.* 2016; 1415:91-105.

[26] Finn RD, Attwood TK, Babbitt PC, et al. InterPro in 2017—beyond protein family and domain annotations. *Nucleic Acids Res.* 2017;45:190–199.

[27] Kozakov D, Hall DR, Xia B, et al. The ClusPro web server for protein-protein docking. *Nat Protoc.* 2017;12:255–278.

[28] Zhu X, Mitchell JC. KFC2: a knowledge-based hot spot prediction method based on interface solvation, atomic density, and plasticity features. *Proteins.* 2011;79:2671–2683.

[29] Zhang Y, Morar M, Ealick SE. Structural biology of the purine biosynthetic pathway. *Cell Mol Life Sci.* 2008;65:3699-724.

[30] Cauët E, Rooman M, Wintjens R, Liévin J, Biot C. Histidine−Aromatic Interactions in Proteins and Protein−Ligand Complexes: Quantum Chemical Study of X-ray and Model Structures. *J Chem Theory Comput.* 2005;1:472–483.

[31] Plumbridge JA. Repression and induction of the nag regulon of *Escherichia coli* K-12: the roles of NagC and NagA in maintenance of the uninduced state. *Mol Microbiol.* 1991;5:2053-62.

[32] Plumbridge J. Co-ordinated regulation of amino sugar biosynthesis and degradation: the NagC repressor acts as both an activator and a repressor for the transcription of the glmUS operon and requires two separated NagC binding sites. *EMBO J.* 1995;14:3958-65.

[33] Finn RD, Attwood TK, Babbitt PC, Bateman A, Bork P, Bridge AJ, et al. InterPro in 2017-beyond protein family and domain annotations. *Nuc Acid Res.* 2017;45:D190-D9.

[34] Murzin AG. Structure classification-based assessment of CASP3 predictions for the fold recognition targets. *Proteins.* 1999;Suppl 3:88-103.

[35] Akira T, Komatsu M, Nango R, Tomooka A, Konaka K, Yamauchi M, et al. Molecular cloning and expression of a rat cDNA encoding 5-aminoimidazole-4-carboxamide ribonucleotide formyltransferase/IMP cyclohydrolase. *Gene.* 1997;197:289-93.

[36] Staal FJT, van der Burg M, Wessels LFA, et al. DNA microarrays for comparison of gene expression profiles between diagnosis and relapse in precursor-B acute lymphoblastic leukemia: choice of technique and purification influence the identification of potential diagnostic markers. *Leukemia.* 2003;17:1324–1332.

[37] Silva K. Hot spots and single nucleotide polymorphisms on the interaction interface of RAD51 and p53 complex. *J Tre Bio Res.* 2018;1:1–5.

[38] Tannous I, Santos T, de Curcio J, et al. Involvement of protein-protein interactions of eNOS and genetic polymorphisms in coronary artery disease. *Int J Clin Cardiol Res.* 2018;02:067–071.

[39] de Curcio JS, Lima RM, Oliveira, LN, et al. Structure-based design of TFF3-PAR2 inhibitor peptides as a promising new therapeutic approach for endometriosis patients. *M J Gyne.* 2018;03:06.

[40] Li Y, Huang Y, Swaminathan CP, Smith-Gill SJ, Mariuzza RA. Magnitude of the hydrophobic effect at central versus peripheral sites in protein-protein interfaces. *Structure.* 2005;13:297–307.

[41] Norel R, Sheinerman F, Petrey D, Honig B. Electrostatic contributions to protein–protein interactions: Fast energetic filters for docking and their physical basis. *Protein Sci.* 2001;10:2147–2161.

[42] Zhang Z, Witham S, Alexov E. On the role of electrostatics on protein-protein interactions. *Phys Biol.* 2011;8:035001.

[43] Li L, Jia Z, Peng Y, et al. Forces and Disease: Electrostatic force differences caused by mutations in kinesin motor domains can distinguish between disease-causing and non-disease-causing mutations. *Sci Rep.* 2017;7:8237.

[44] Feng J, Hu Y, Yi H, et al. Two conserved arginine residues from the SK3 potassium channel outer vestibule control selectivity of recognition by scorpion toxins. *J Biol Chem.* 2013;288:12544–12553.

[45] Tesei G, Vazdar M, Jensen MR, et al. Self-association of a highly charged arginine-rich cell-penetrating peptide. *Proc Nat Acad Sci USA.* 2017;114:11428–11433.

[46] Dong J-Y, Qin L-Q, Zhang Z, et al. Effect of oral L-arginine supplementation on blood pressure: a meta-analysis of randomized, double-blind, placebo-controlled trials. *Am Heart J.* 2011;162:959–965.

[47] Liao S-M, Du Q-S, Meng J-Z, Pang Z-W, Huang R-B. The multiple roles of histidine in protein interactions. *Chem Cent J.* 2013;7:44.

[48] Goncheva MI, Flannagan RS, Sterling BE, Laakso HA, Friedrich NC, Kaiser JC, et al. Stress-induced inactivation of the *Staphylococcus aureus* purine biosynthesis repressor leads to hypervirulence. *Nat Commun.* 2019;10:775.

[49] Srere PA. The metabolon. *Trends Biochem Sci.* 1985;10:2.

[50] Plumbridge JA. Repression and induction of the *nag* regulon of *Escherichia coli* K-12: the roles of NagC and NagA in maintenance of the uninduced state. *Mol Microbiol.* 1991;5:2053-62.

[51] Maira-Litran T, Kropec A, Abeygunawardana C, Joyce J, Mark G, 3rd, Goldmann DA, et al. Immunochemical properties of the staphylococcal poly-N-acetylglucosamine surface polysaccharide. *Infect Immun.* 2002;70:4433-40.

[52] Tojo M, Yamashita N, Goldmann DA, Pier GB. Isolation and characterization of a capsular polysaccharide adhesin from *Staphylococcus epidermidi*s. *J Infect Dis.* 1988;157:713-22.

[53] Barnhart MM, Lynem J, Chapman MR. GlcNAc-6P levels modulate the expression of curli fibers by *Escherichia coli. J Bacteriol.* 2006;188:5212-9.

[54] Mwangi MM, Wu SW, Zhou Y, Sieradzki K, de Lencastre H, Richardson P, et al. Tracking the *in vivo* evolution of multidrug resistance in *Staphylococcus aureus* by whole-genome sequencing. *Proc Nat Acad Sci USA.* 2007;104:9451-6.

[55] Hoekstra HE, Hirschmann RJ, Bundey RA, Insel PA, Crossland JP. A single amino acid mutation contributes to adaptive beach mouse color pattern. Science (New York, NY. 2006;313:101-4.

[56] Kryazhimskiy S, Plotkin JB. The population genetics of dN/dS. *PLoS Genet.* 2008;4:e1000304.

[57] Spielman SJ, Wilke CO. The relationship between dN/dS and scaled selection coefficients. *Mol Biol Evol.* 2015;32:1097-108.

[58] Nunney L, Schuenzel EL. Detecting natural selection at the molecular level: a reexamination of some "classic" examples of adaptive evolution. *J Mol Evol.* 2006;62:176-95.

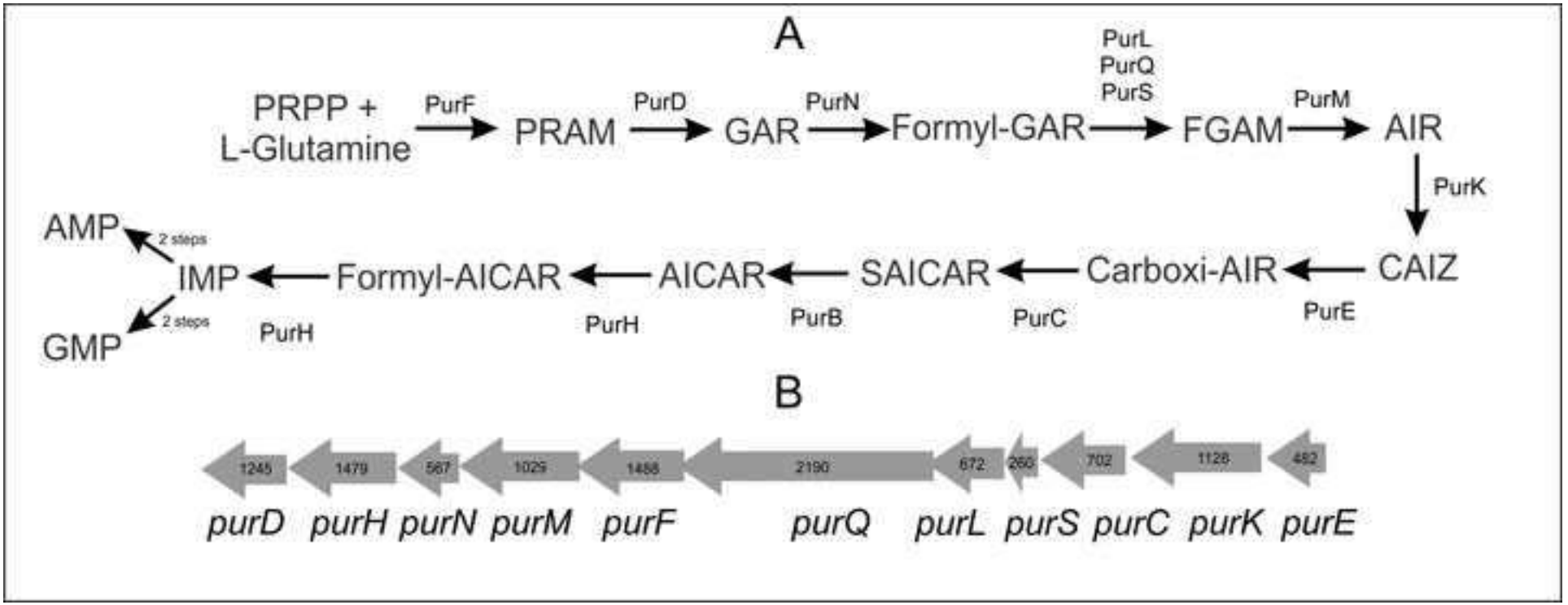


Figure 2

Click here to access/download;Figure;Figure 2.jpg

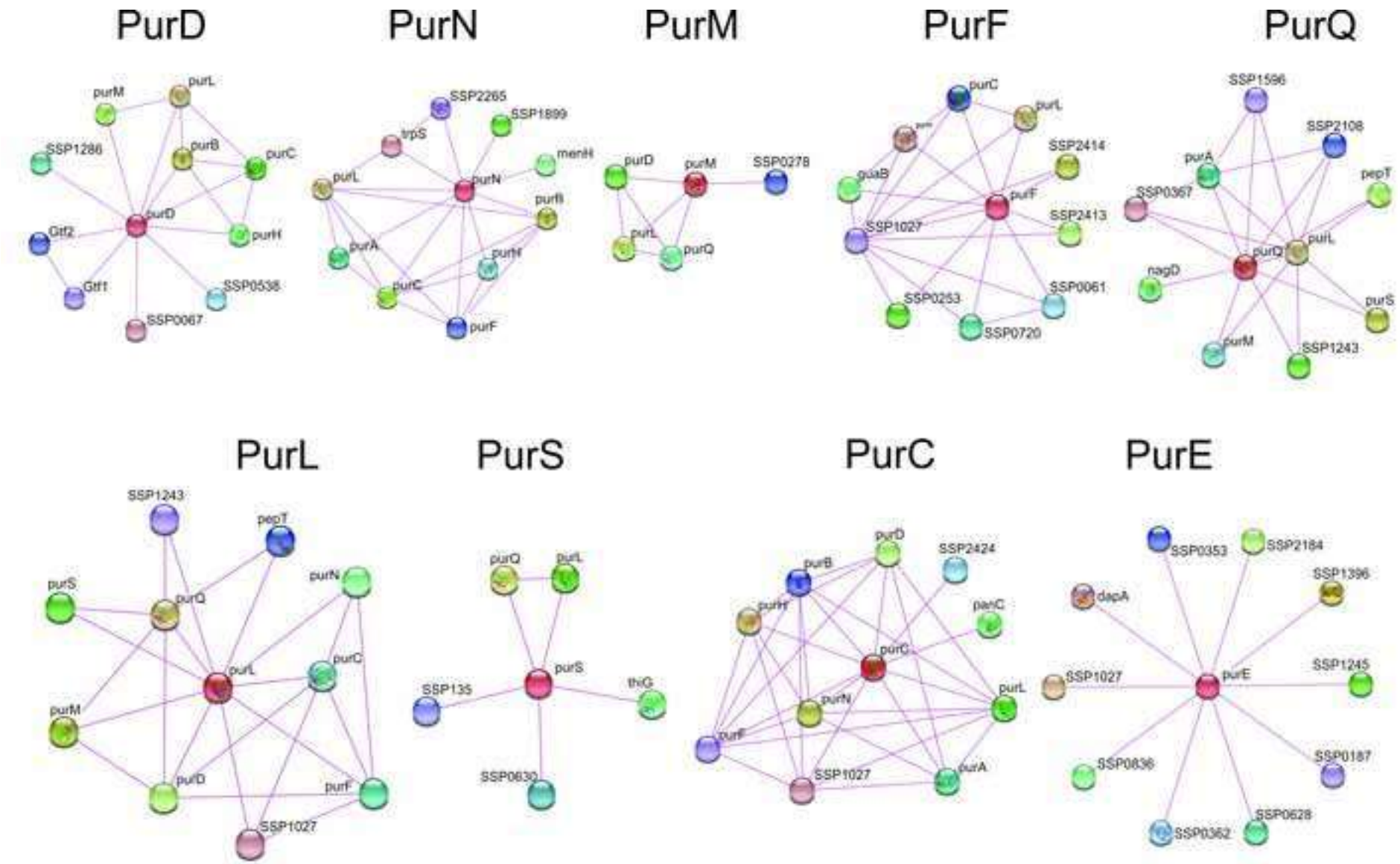


Figure 3

Click here to access/download;Figure;Figure 3.jpg

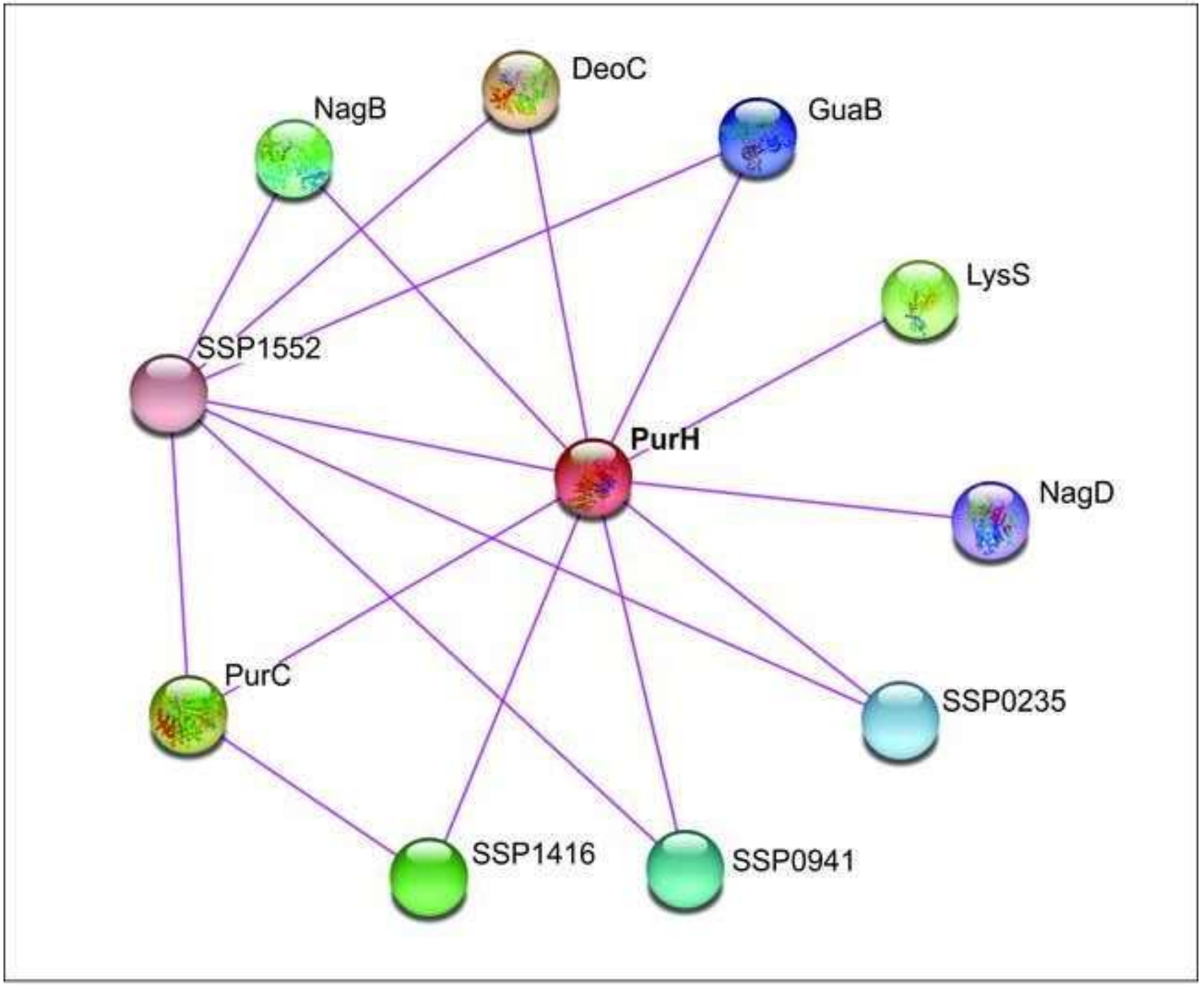


Figure 4

Click here to access/download;Figure;Figure 4.jpg

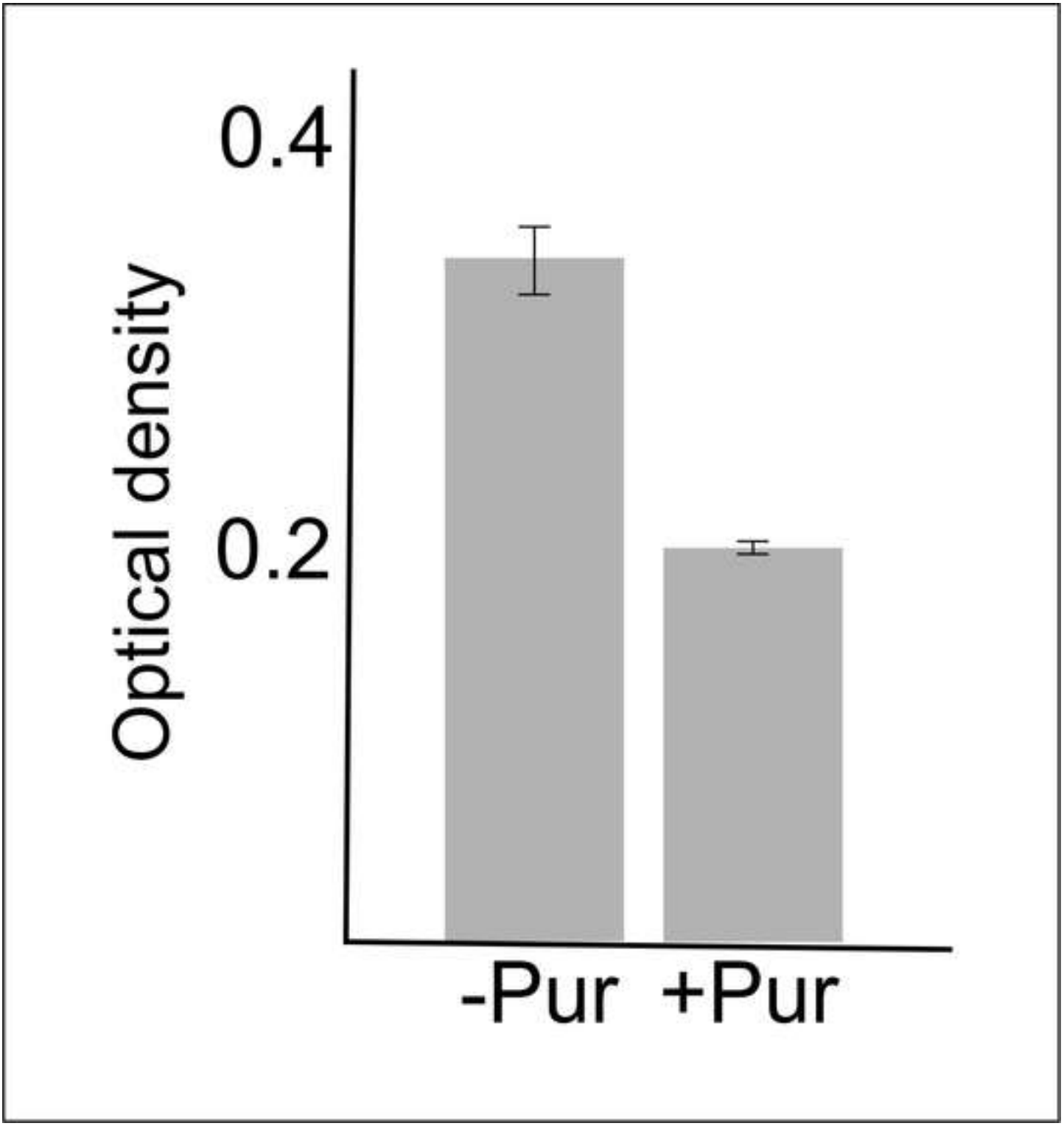


Figure 5

Click here to access/download;Figure;Figure 5.jpg

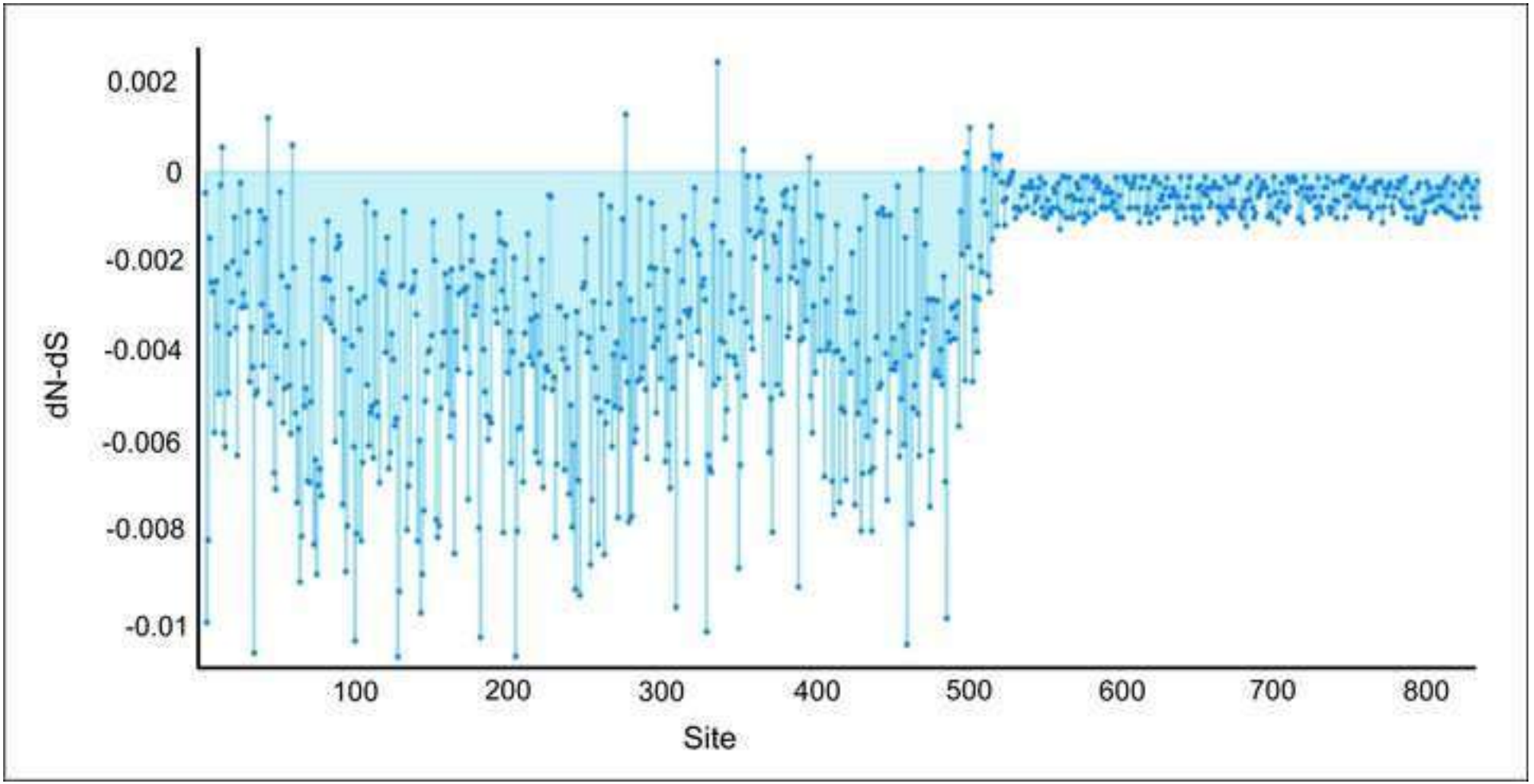


Figure 6

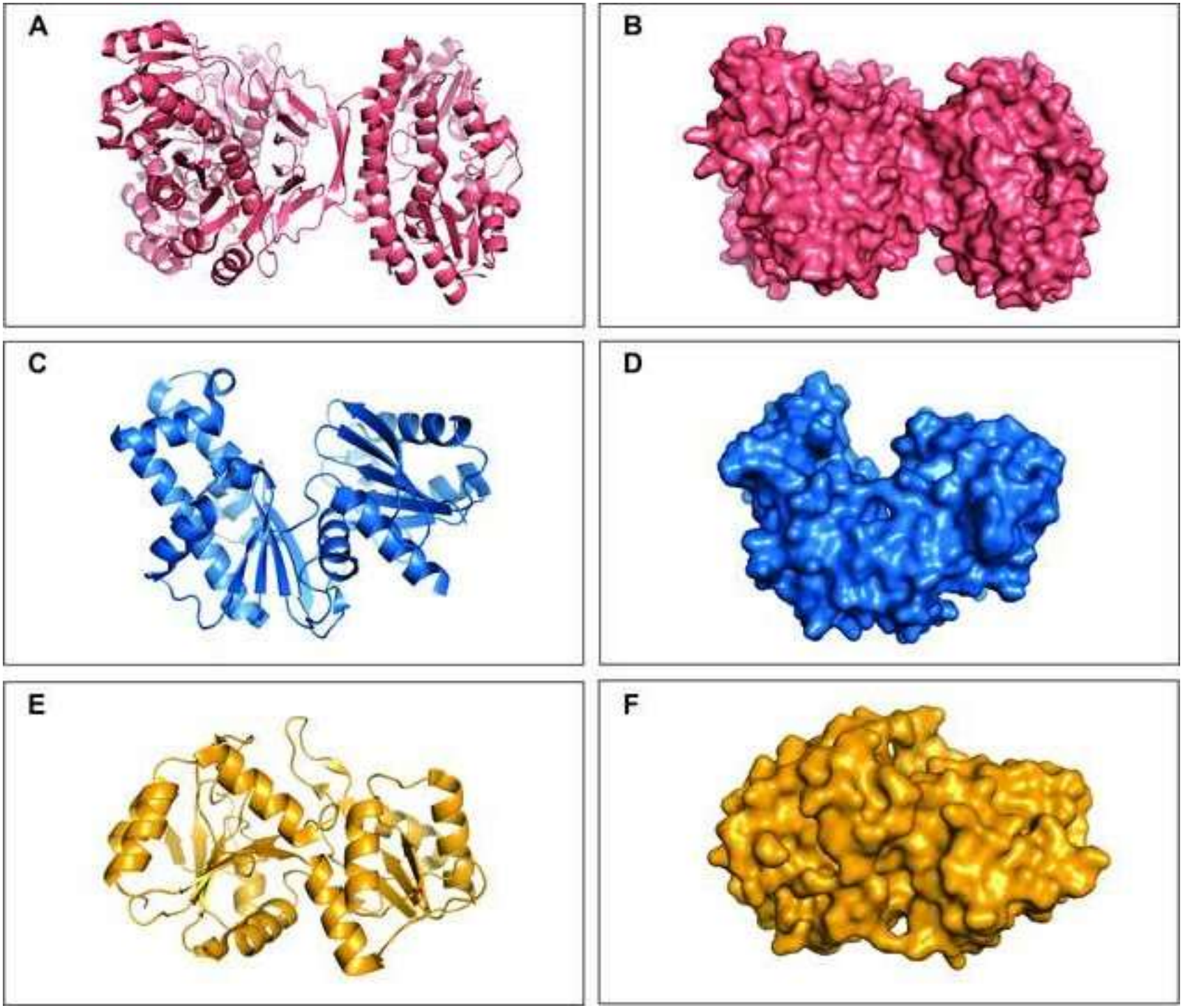


Figure 7

Click here to access/download;Figure;Figure 7.jpg

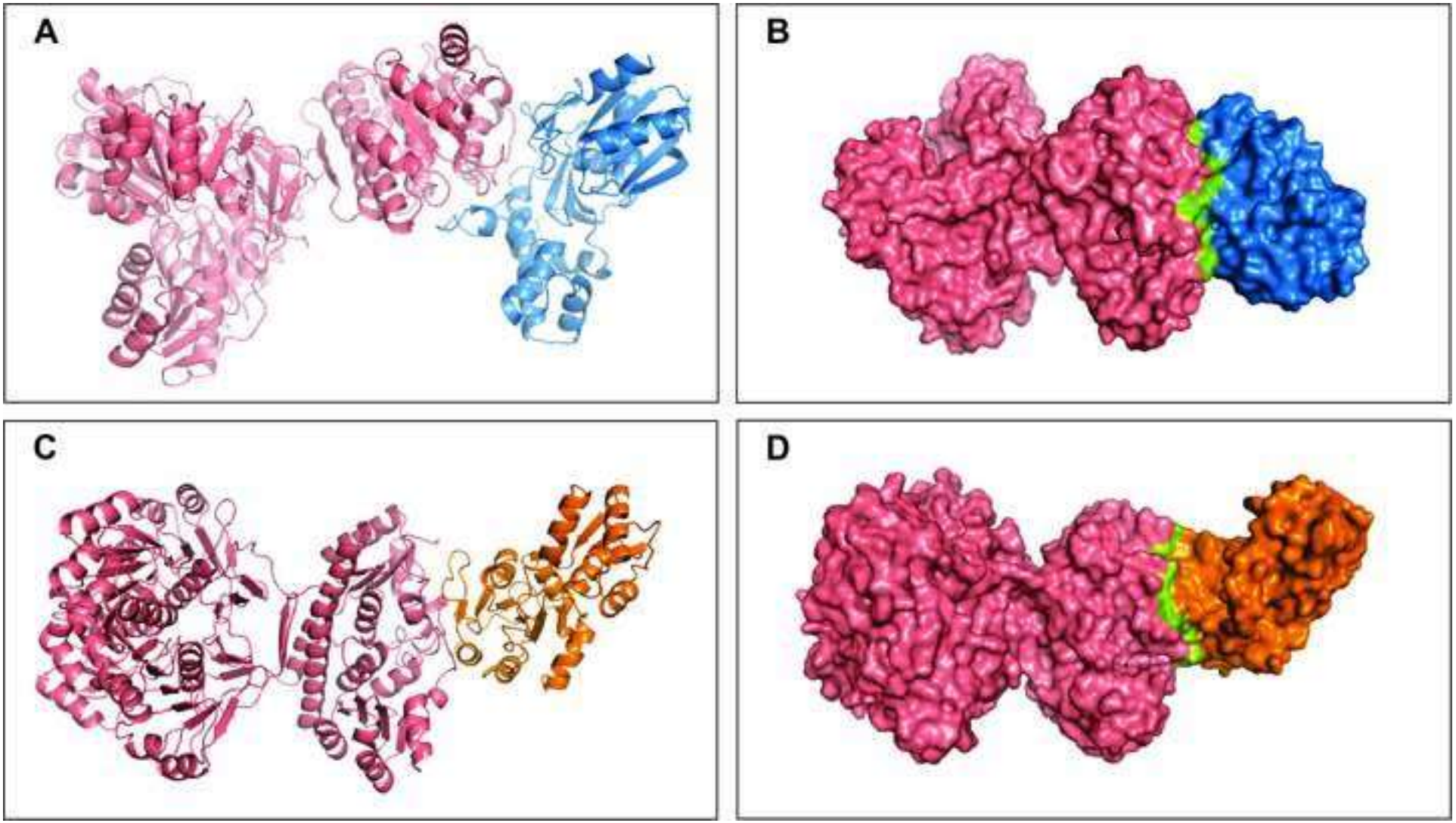
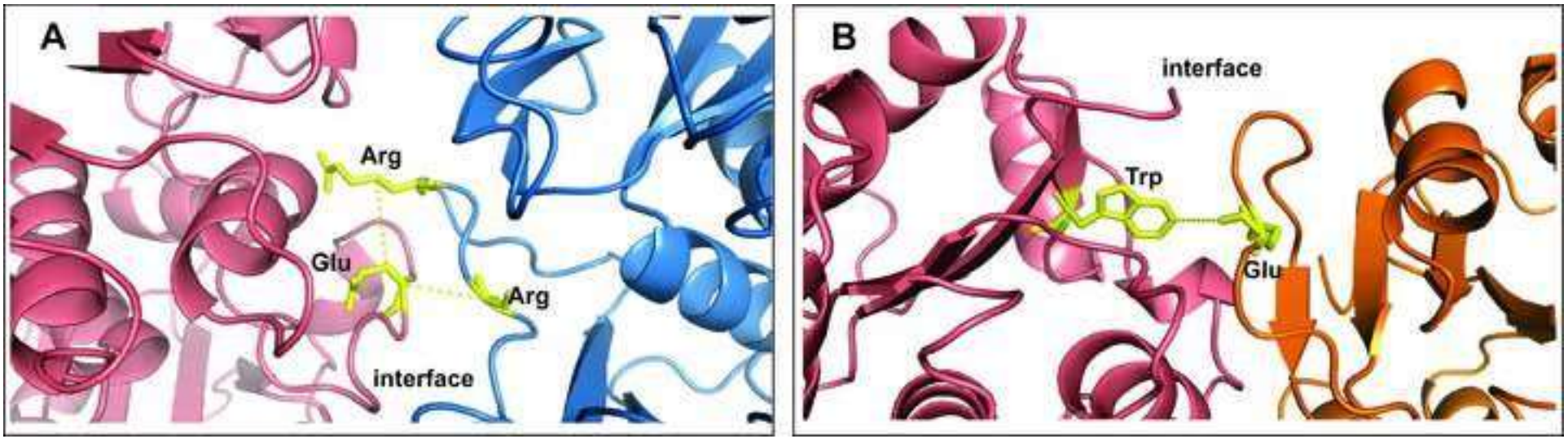


Figure 8

Click here to access/download;Figure;Figure 8.jpg



Supplementary Figure 1

Supplementary Table 1

Declaration of interests

☒The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: