UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE FÍSICA DE SÃO CARLOS

ÉVERTON EDÉSIO DINIS SILVA

ESTUDOS SOBRE A REGULAÇÃO DOS TRANSDUTORES DE
SINAL PA0847 E HSBD DE PSEUDOMONAS AERUGINOSA E
SOBRE SISTEMAS GGDEF E EAL EM TANDEM

São Carlos
2018

ÉVERTON EDÉSIO DINIS SILVA

Estudos sobre a regulação dos transdutores de sinal PA0847 e HsbD
de *Pseudomonas aeruginosa* e sobre sistemas GGDEF e EAL em
tandem

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

Área de concentração: Física Aplicada
Opção: Física Biomolecular
Orientador: Prof. Dr. Marcos Vicente de
Albuquerque Salles Navarro

Versão Corrigida

(Versão original disponível na Unidade que aloja o Programa)

São Carlos
2018

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

Silva, Éverton Edésio Dinis Silva

Estudos sobre a regulação dos transdutores de sinal PA0847 e HsbD de *Pseudomonas aeruginosa* e sobre sistemas GGDEF e EAL em tandem. / Éverton Edésio Dinis Silva Silva; orientador Marcos Vicente de Albuquerque Salles Navarro Navarro - versão corrigida -- São Carlos, 2018. 256 p.

Tese (Doutorado - Programa de Pós-Graduação em Física Aplicada Biomolecular) -- Instituto de Física de São Carlos, Universidade de São Paulo, 2018.

1. c-di-GMP. 2. Sinalização bacteriana. 3. *Pseudomonas aeruginosa*. I. Navarro, Marcos Vicente de Albuquerque Salles Navarro, orient. II. Título.

Dedico aos meus familiares, aos grandes amigos e a você leitor, porque
ciência é também sobre compartilhar.

AGRADECIMENTOS

Ao Instituto de Física de São Carlos, pela excelente infraestrutura e suporte técnico-administrativo.

À Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), pelo apoio financeiro imprescindível com a bolsa de doutorado direto, pelo processo 2012/25276-6.

Ao Prof. Marcos Navarro, pela oportunidade desde a iniciação científica e por todos os ensinamentos.

Ao LNLS (CNPEN), que cederam as linhas de luz MX2 para os experimentos e forneceram excelente suporte técnico.

A todos os meus familiares, em especial aos meus pais, Maria Fátima Oliveira Silva e Délcio Arlindo da Silva e minhas irmãs Elizett Santos (Zett), Elisângela Silva (Zanja), Jennifer Oliveira (Deive) e Eliana Silva (Nana). Mesmo distantes, me apoiaram, compreenderam minhas faltas e sempre foram meu porto seguro.

A todo corpo técnico do laboratório de biofísica Sérgio Mascarenhas, principalmente a Andressa Pinto (Drê), Fernando Lima (Fer), Isabel Moraes (Bel), Júlia Martins e Rafael Panhota (Rafa). Vocês foram muito mais que técnicos, foram super-heróis, sempre em prontidão para o que eu precisava, mesmo para assuntos pessoais. Muito do doutorado devo a vocês!

Agradeço também aos amigos e companheiros de orientação e de laboratório no grupo de Biofísica: Ana Eliza Zeraik, Ana Felizatti (Aninha), Andrea Cardoso (Déia), Bruno Matsuyama, Carla Martins, Danielle Castro (Dani), Débora Órcia, Diego Sastre, Edson Crusca, Flávio Rosseto, Gabriel Belém (Gabs), Gabriel Brognara (Netinho), Heline Helen, Heloisa Ciol (Helo), Helton Wiggers, Jakeline Ferreira (Jack), Joci Macêdo, Juliana Cheleski (Ju), Leonardo Delphito (Leo), Luis Basso (Militar), Marcela Argentin (Marcelinha), Naiara Torres (Nai), Nathalya Mesquita (Naty), Patrícia Kumagai (Paty), Raphael Meneghelo (Rapha), Samuel Guimarães, Sinara Brasil (Sina), Sumária Souza (Suma) e Wesley Generoso. Todos colaboraram com esse trabalho de alguma forma, direta ou indiretamente, assim como colaboraram no meu crescimento pessoal e científico. Além disso, vocês proporcionaram momentos de descontração ótimos nos intervalos e tornaram a caminhada muito mais agradável.

Gostaria de agradecer também aos professores Ana Paula Araújo e Richard Garratt pelo apoio, especialmente na reta final do doutorado. Agradecer ainda aos técnicos do laboratório vizinho de Cristalografia, Bianca Saes, Humberto Pereira e Susana Andrea, pelas diversas vezes

que quebrarem o galho. Gostaria de agradecer ao Ivan Rosa pelas discussões e incentivos e também a Paula Lins e ao Prof. Júlio Borges, que auxiliaram em experimentos de ITC.

Agradeço a todos os amigos da graduação em Física Biomolecular (Biomol), em especial cito Anacleto Sousa, Bruno Ono (Kuririm), Érika Chang (Keka), Mariana Koyama (Mary), Naiara Torres (Nai), Paola Lanzone (Beta) e Raissa Gutierrez (Ra). E agradecer também aos grandes amigos de jornada, que conheci no alojamento da USP São Carlos (aloja) e na Republica CC. Especialmente agradecer a Afonso Vaz (Fon), Eduardo Martins (Próximo), Gabriel Santos, Felipe Ceregatti (Lipe), Leandro Viscardi (Mestre), Matheus Pereira (Treta) e Rafael Ceregatti (Rafa). Vocês foram ótimos companheiros, me receberam de braços abertos, abriram meus olhos para novas perspectivas e me mostraram que tenho muito ainda para aprender. A amizade de vocês me deu forças para continuar. Vou levá-los sempre comigo!

"The first principle is that you must not fool yourself and you are the easiest person to fool"

Richard Feynman

RESUMO

SILVA, E. E. D. **Estudos sobre a regulação dos transdutores de sinal PA0847 e HsbD de *Pseudomonas aeruginosa* e sobre sistemas GGDEF e EAL em *tandem***. 2018. 256 p. Tese (Doutorado em Ciências) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2018.

Proteínas sinalizadoras membranares são amplamente usadas por bactérias para adequar seu metabolismo a estímulos ambientais. Elas possuem ainda um grande potencial de aplicação, como em técnicas para controle de infecções e o desenvolvimento de ferramentas biotecnológicas, como biossensores. Entretanto, pouco é conhecido sobre os mecanismos estruturais de transdução bacterianos. Em particular, pouco é descrito sobre proteínas sinalizadoras que possuem os domínios GGDEF e/ou EAL, responsáveis por desencadear a síntese, degradação ou captação da molécula c-di-GMP, um dos principais reguladores do metabolismo bacteriano. Por isso, esse estudo objetivou a caracterização de duas proteínas sinalizadoras transmembrana de *Pseudomonas aeruginosa*, codificadas nos genes PA0847 e PA3343 (HsbD), além de trazer análises por bioinformática de proteínas com domínios GGDEF e EAL consecutivos, comumente associados a membrana celular. É relatado que PA0847 e HsbD são ambas diguanilato ciclases (DGC), que impactam consideravelmente em fenótipos de *P. aeruginosa*, patógeno modelo de estudo e que infecta pacientes imunocomprometidos. Além disso, um estudo anterior mostrou a interação entre a porção citoplasmática de PA0847 e a região N-terminal de HsbD (HsbD-Nt), pela técnica de duplo híbrido. Nesse estudo, é mostrado que essa interação, associada à presença de um padrão conservado em HsbD-Nt, sugere a presença um sistema de controle específico de c-di-GMP. Ensaio *in vitro* de pull-down, termoforese e duplo híbrido, utilizando domínios isolados de PA0847, confirmaram a interação com HsbD-Nt e mostraram que a interação é dependente de todos os domínios de PA0847. Análises de massa molecular e de cinética enzimática mostraram que o HAMP eleva a atividade DGC e promove a formação de tetrâmeros. A presença de um transportador de sulfato compartilhando um operon com PA0847, associado a modelos computacionais da região periplasmática de PA0847, suportam o envolvimento de PA0847 na homeostase desse íon. Nesse estudo foram ainda identificados, por meio de análises de sequências, novos potenciais grupos de resíduos que estariam envolvidos na regulação de proteínas GGDEF-EAL. Análises adicionais de sequências e de estruturas indicaram também um mecanismo de dimerização alternativo para EALs, quando a região de dimerização está ausente. Estes

resultados são inéditos e fornecem passos para uma melhor compreensão da sinalização bacteriana intermediada pelo c-di-GMP.

Palavras-chave: c-di-GMP. Sinalização bacteriana. *Pseudomonas aeruginosa*.

ABSTRACT

SILVA, E. E. D. **Regulation studies of signalling transducers PA0847 and HsbD from *Pseudomonas aeruginosa* and GGDEF and EAL systems in tandem.** 2018. 256 p. Tese (Doutorado em Ciências) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2018.

Membrane signaling protein are widely used by bacteria to adequate their metabolism of environmental cues. They present a great potential of applicability, such as techniques to control bacterial infections and development of biotechnological devices, for example in biosensors. However, little is known regarding the mechanisms of signal transduction in bacteria. In particular, about receptors that posses GGDEF and/or EAL domains, performing synthesis, degradation or recognition of c-di-GMP, which is one of the major bacteria metabolism regulators. For this reason, the present study aimed characterize two important transmembrane signaling proteins from *Pseudomonas aeruginosa*, coded in genes PA0847 and PA3343 (HsbD) and performed bioinformatic analysis of proteins with GGDEF and EAL in tandem, usually associated with membrane. Previous studies indicated that PA0847 and HsbD are both diguanylate cyclases (DGCs) that significantly impair phenotypes of *P. aeruginosa*, which is a model organism and common cause of infection in immunocompromised patients. In the present study we have shown that this interaction associated with the presence of a well-conserved pattern from HsbD-Nt, suggests the presence a system used for specific downregulation of c-di-GMP. Furthermore, using isolated domains of PA0847 in thermophoresis, pull down and two-hybrid confirmed the interaction and showed that the full PA0847 cytoplasmic is necessary to interaction with HsbD. Additional amalysis of molecular mass and enzyme kinetic showed that HAMP increases DGC activity, alters the co-purification pattern of c-di-GMP and promotes tetramers formation in PA0847 soluble constructs. The presence of sulfate transporter in a common operon with PA0847, associated with computation model of periplasmatic region of PA0847 supports the relation of this protein in sulfate homeostasis. We also identified using new residues groups copled that are potential play roles in GGDEF-EAL proteins. We found by stuctural analisys alternative mechanisms of dimerization in EAL domais that do not posses dimerization site conserved. These results have not been presented elsewhere and provide insights into the bacterial signaling of c-di-GMP pathways.

Keywords: c-di-GMP. Bacterial signaling. *Pseudomonas aeruginosa*.

LISTA DE FIGURAS

Figura 1 - Esquematização das etapas do ciclo do biofilme bacteriano e da localização das proteínas em estudo.	33
Figura 2 - Exemplificações de danos causados por biofilmes.....	35
Figura 3 - Esquematização dos dispositivos médicos susceptíveis ao crescimento de biofilme e patologias agravadas e/ou causadas por essa formação microbiana	36
Figura 4 - Microrganismos dos quais as proteínas foram abordadas no presente estudo..	37
Figura 5 - Estruturas moleculares de nucleotídeos usados encontrados como segundos mensageiros clássicos.	41
Figura 6 - Arquitetura e equilíbrio dos estados oligoméricos apresentado em solução do c-di-GMP.	42
Figura 7 - Esquematização de exemplos das vias de sinalização do c-di-GMP.....	43
Figura 8 - Mecanismos propostos para o direcionamento do c-di-GMP para receptores e experimento que corrobora para o modelo específico (<i>local pool</i>).	45
Figura 9 - Estrutura característica do motivo GGDEF.....	47
Figura 10 - Distribuição dos elementos do GGDEF sequencial e tridimensional.....	49
Figura 11 - Representação das estruturas dos motivos GGDEF e DXLT, característicos de domínios GGDEF.	50
Figura 12 - Dímero do domínio GGDEF na conformação ativa exemplificada pelo sensor de zinco DgcZ.....	51
Figura 13 - Estrutura dois inibidas de DGCs, ilustrados pelas estruturas de PleD de PDB 2v0n e 1w25.....	52
Figura 14 - Modos de dimerização em DGCs e receptores promovidos pela ligação do c-di-GMP dimérico.	53
Figura 15 - Distribuição dos elementos do EAL sequencial e tridimensional.	57
Figura 16 - Visão superior do dímero de EALs, com destaque para o eixo de simetria (TBD1265, PDB:3n3t).....	60
Figura 17 - Estruturas cristalográficas de YahA, com e sem o ligante, e BlrP1 com o ligante.	62
Figura 18 - Resumo de arquiteturas de domínios de algumas das proteínas citadas nesse texto, grande parte detentoras tanto do domínio GGDEF quanto do EAL.....	63

Figura 19 - Esquematisação do mecanismo de atuação da LapD.....	64
Figura 20 - Estruturas dos dímeros de GGDEF-EAL de STM3615 (resolvida pelo nosso grupo) e MorA. ¹¹⁶	64
Figura 21 - Ilustração da variedade de possíveis combinações de domínios de vias de sinalização bacterianas (A) e algumas das proteínas das vias do c-di-GMP de <i>P. aeruginosa</i> (B), exemplificando essa variedade.....	66
Figura 22 - Analogia entre sistema de sinalização clássico MCPs (A) e outro que inclui uma DGC (B).	68
Figura 23 - Modelo ilustrativo geral dos componentes básicos que podem aparecer em sistemas de transdução de sinal abordados no trabalho.	69
Figura 24 - Divisão e diversidade dos sistemas conhecidos de recepção de sinal ambiental.	70
Figura 25 - Enovelamento PAS.	73
Figura 26 - Estruturas RbdA (A) e VicK (B), exemplo nos quais o PAS é empregado para dimerização.	74
Figura 27 - Distribuição de domínios das vias do c-di-GMP nos grupos CHASE.....	76
Figura 28 - Exemplificação das arquiteturas secundárias e terciárias de domínios CHASE, ilustrados pelos CHASE e CHASE4 (enovelamento PDC) e pelo CHASE3.	77
Figura 29 - Elementos do domínio HAMP, sequencial e tridimensional (A) e (B), e os mecanismos descritos para os domínios em (C), (D) e (E).	78
Figura 30 - Exemplos da transmissão de sinal pela hélice C-terminal α do domínio LOV, que possui enovelamento PAS (A) e padrão hidrofóbico <i>coiled-coil</i> comumente conectando PAS e GGDEFs (B).....	81
Figura 31 - Tipos de mudanças conformacionais transmitidas por porções sensoras propostas.	82
Figura 32 - Arquitetura ilustrativa de PA0847, das proteínas que possuem interação com a fração citossólica da mesma (PA0847-HPG-1) e de STM3615.....	84
Figura 33 - Reprodução de alguns dos resultados de análise de biofilme das cepas com superexpressão (marcados com ++) e deleção (marcados com -) de genes envolvidos no presente trabalho. ¹⁹³	85
Figura 34 - Esquematisação de dados da literatura sobre a proteína HsbD.....	86
Figura 35 - Arquitetura ilustrativa da distribuição de domínios de STM3615.	87

Figura 36 - Esquematização simplificada dos passos de purificação com protocolo padrão.	100
Figura 37 - Exemplificação do cromatogramas de purificação do c-di-GMP.....	104
Figura 38 - Gradiente utilizado nas análises de c-di-GMP (A) e a resultado da análise de atividade de STM3615 ²⁴²⁻⁶⁶⁸ utilizando o método apresentado(B).....	106
Figura 39 - Resultados de massas das amostras purificadas de c-di-GMP.	107
Figura 40 – Montagem experimental do SEC-MALS.....	111
Figura 41 - Esquematização ilustrativa dos processos para medição com <i>kit</i> comercial EnzChek.....	112
Figura 42 - Esquematização simplificada da montagem experimental do MST.....	116
Figura 43 - Esquematização simplificada do experimento de duplo híbrido bacteriano utilizado.	117
Figura 44 - Ilustração experimental do <i>pull-down</i>	119
Figura 45 - Exemplificação do planejamento de múltiplas construções aplicados para o domínio CHASE4 de PA0847.....	130
Figura 46 - Construções de PA0847 e o padrão do impacto da remoção das porções do domínio HAMP em proteínas inteiras de Tsr, sensor quimiotático de serina.	131
Figura 47 - Parte do alinhamento entre GGDEFs das proteínas em estudo com outras de referência.	132
Figura 48 - Resumo detalhado das arquiteturas de domínios e das construções 13 construções de PA0847 (A) e da construção da porção N-terminal de HsbD (B).	132
Figura 49 - Análise de restrição em gel de agarose das construções de PA0847 e PA3343 em pETSUMO.....	134
Figura 50 - Análise em gel de agarose de amostras das PCRs para obtenção de pETSUMO-HPG-1 ^{AXXD}	134
Figura 51 - Resultados das etapas de purificação aplicadas para construções da PA0847.	135
Figura 52 - Resultado do SDS-PAGE 15 % das amostras finais das construções solúveis de PA0847.	135

Figura 53 - Cromatogramas de amostras da HPG-1, ilustrando o equilíbrio oligomérico (A) e o ganho de resolução com diferentes colunas de exclusão molecular (B).	137
Figura 54 - Gel de análise de restrição de HsbD em pETSUMO.	138
Figura 55 - Análises em SDS-PAGE e cromatograma final das etapas de purificação de HsbD-Nt.	139
Figura 56 - Cromatogramas do HPLC dos extratos das construções para análise de c-di-GMP copurificado.	142
Figura 57 - Cromatogramas de purificações em Superdex 200 10/300 (GE) de construções de PA0847, com e sem tratamento PDE.	144
Figura 58 - Curva padrão de massas para gel filtração, utilizando a coluna Superdex200 10/300 (GE).	145
Figura 59 - Cromatogramas de exclusão molecular em Superdex200 10/300 (GE), usados para cálculo de massas.	146
Figura 60 - Cromatogramas produzidos em experimentos de SEC-MALS das construções de PA0847.	147
Figura 61 - Alinhamento entre o HAMP de PA0847 e outros exemplares do domínio já estudados.	150
Figura 62 - Análises de SEC-MALS dos picos oligoméricos da Superdex 200 10/300 (GE).	151
Figura 63 - Resultados de SEC-MALS de construções da PA0847 com e sem c-di-GMP.	153
Figura 64 - Sobreposição dos perfis de eluição no experimento de SEC-MALS de construções da PA0847 com diferentes concentrações de NaCl na amostra aplicada. Coluna utilizada: STR-C SEC-150 (Sepax).	155
Figura 65 - Análises de HPLC de padrões e de amostras de reações com GTP (A) e de reações com construções da PA0847 (B).	156
Figura 66 - Resultados de cinética das construções de PA0847, com e sem NaCl adicionado.	157
Figura 67 - Gráficos de atividade das construções de PA0847, com e sem a adição de NaCl no meio reacional.	159
Figura 68 - Cristais da construção periplasmática da PA0847 C4-2.	161
Figura 69 - Mecanismos propostos para proteínas com enovelamento <i>double</i> -PDC, também identificado para porção periplasmática de PA0847.	163

Figura 70 - Predição de operon de PA0847 e PA0846, disponibilizado pelo servidor pseudomonas.com.....	164
Figura 71 - Parte do alinhamento de alguns domínios PAS conhecidos (B) e árvore filogenética a construída a partir do mesmo (B).....	167
Figura 72 - Experimento de MST utilizando HP e P marcas com o fluoróforo e variando a concentração da construção G.	168
Figura 73 - Agrupamento de sequências resultantes de acordo com a arquitetura de domínios (A) e padrão do alinhamento das sequências recuperadas (B). ...	169
Figura 74 – <i>Pull- down</i> imobilizando HsbD-Nt fusionada a SUMO-6xHis (SUMO-Nt) e recuperando a HPG-1.	170
Figura 75 - Resultados de crescimento celular dos experimentos de duplo híbrido bacteriano utilizando diferentes construções de PA0847 contra os N-terminais de HsbD e PA1851 (HsbD-Nt e PA1851-Nt).	171
Figura 76 - Curvas de MST utilizando HPG-1 marcada e variando a concentração de amostras de HsbD-Nt.....	173
Figura 77 - Curvas de interação entre construções PA0847 marcada com NT-647-NHS e adicionando diferentes concentrações de HsbD-Nt (pico 3 da SEC).	174
Figura 78 - Teste de atividade de HPG-1 na presença de diferentes concentrações de 6xHis-SUMO-HsbD-Nt (SUMO-Nt).	175
Figura 79 - Representação das estruturas do domínio GGDEF na conformação ativa e inibida (acima) e das representações simplificadas (abaixo).....	176
Figura 80 - Formas diméricas possíveis das construções G, PG e HPG-2 ou HPG-3.....	177
Figura 81 - Esquematisações dos equilíbrios possíveis de dímeros e tetrâmeros para os domínios estudados de PA0847.....	178
Figura 82 - Mecanismo para PA0847 para formação de tetrâmeros (A) e sua alternância entre estados ativos e inativos (B).	179
Figura 83 - Mecanismo ilustrativo da sinalização para PA0847 e HsbD.....	180
Figura 84 - Distribuição das sequências recuperadas divididos em relação a conservação do sítio inibitório RXXD.	183
Figura 85 - Rede de correlação e divisão das comunidades de acordo com os grupos de conservação dos domínios GGDEF e EAL.	186
Figura 86 - Análises estrutural das comunidades obtidas do programa PFstats.....	188
Figura 87 - Região do alinhamento envolvida na dimerização de alguns domínios EALs.	192

Figura 88 - Estruturas das regiões de dimerização de alguns domínios EALs com estrutura resolvida.	193
Figura 89 - Esquematização das proteínas e respectivas construções que foram abordadas no trabalho.	224
Figura 90 - Alinhamento de diversos domínios GGDEF de proteínas já bem estudadas na literatura.	231
Figura 91 - Alinhamento de diversos domínios EALs de proteínas abordadas na literatura e de proteínas.	232
Figura 92 - Alinhamento de domínios PAS com responsáveis por captar diferentes estímulos e que são bem caracterizadas.	233
Figura 93 - Esquema simplificado dos passos de clonagem utilizando enzimas de restrição.	240
Figura 94 - Gráfico ilustrando o cálculo de massa molecular e raio de giro médio, através do experimento de SEC-MALS.	244
Figura 95 - Ilustração do processamento de dados para o cálculo do raio hidrodinâmico (rh).	245
Figura 96 - Esquematização dos procedimentos para cálculo de K_m e $V_{máx}$	248
Figura 97 - Exemplificação ilustrativa das curvas e dos cálculos do experimento de MST para cálculo de K_d	252
Figura 98 - Exemplificação esquemática das análises feitas com o PFstats usando um alinhamento modelo.	256

LISTA DE TABELAS

Tabela 1 - Resumo de resíduos do domínio EAL importantes para a atividade.	56
Tabela 2 - Dados de exemplos do efeito de mutações em resíduos importantes sobre parâmetros cinéticos.....	59
Tabela 3 - Dados da literatura de superexpressão de determinadas DGCs	85
Tabela 4 - Esquematização dos ciclos do termociclador nos quais foram submetidas as PCRs.....	92
Tabela 5 - Resumo esquemático dos ciclos utilizados no termociclador para PCRs de mutação sítio-dirigida.	96
Tabela 6 - Aplicação dos tampões na coluna de fase reversa para purificação de c-di-GMP.	104
Tabela 7 - Aplicação dos tampões na coluna de fase reversa para análise de nucleotídeos.	105
Tabela 8 - Aplicação dos tampões na coluna de fase reversa para análise de atividade DGC.	107
Tabela 9 - Soluções utilizadas para extração de gel e tripsinização para análise em espectrômetro de massas.	108
Tabela 10 - Tampões preparados voltados para busca de um que otimize a estabilidade das construções.....	120
Tabela 11 - Posições e conjunto as opções de resíduos que devem ocupá-las para classificar como conservado os domínios ou o motivo.....	126
Tabela 12 - Detalhes das construções planejadas para trabalhos em solução e alguns de seus respectivos parâmetros.	133
Tabela 13 - Resumo dos resultados nas etapas de produção e rendimento médio das preparações para PA0847.....	136
Tabela 14 - Resumo de resultados para HsbD-Nt (pico 3 da gel filtração).....	140
Tabela 15 - Resultados de oligomerização por SEC analítica (aSEC) e SEC-MALS. ...	148
Tabela 16 - Dados de SEC-MALS dos diferentes picos de eluição da purificação de exclusão das construções HPG-1 e PG	152
Tabela 17 - Resultados de SEC-MALS de comparação dos efeitos da incubação com c-di-GMP.	153

Tabela 18 - Variação das populações oligoméricas frente a concentração de NaCl na amostra aplicada.	155
Tabela 19 - Dados de coleta e de processamento para C4-2 com selenometionina (LNLS).	161
Tabela 20 - Resumo de estruturas utilizadas para construção do modelo da porção periplasmática de PA0847 (resíduos 62-315).....	162
Tabela 21 - Resumo de estruturas utilizadas para construção do modelo da porção PAS de PA0847 (resíduos 402-504).....	166
Tabela 22 - Distribuição das sequências recuperadas nos bancos de dados de acordo com a conservação dos domínios GGDEF e EAL e da conservação do motivo RXXD.....	182
Tabela 23 - Frequência dos resíduos (%) com correlação e variação da frequência em relação aos subgrupos de acordo com a conservação dos domínios GGDEF (abreviação G) e EAL (abreviação E).	185
Tabela 24 - Resumo de resultados de triagens por construções solúveis de diferentes proteínas sensoras transmembranares e de potenciais parceiros de interação.	225
Tabela 25 - Resumo de construções e de primers para produção de proteínas de <i>X. citri</i>	227
Tabela 26 - Resumo de construções e de primers para produção de proteínas de <i>Pseudomonas aeruginosa</i> – de proteínas inteiras em pET29 e construções em pETSUMO.....	228
Tabela 27 - Lista de regiões codificantes que foram otimizadas – adquiridos pela empresa GenScript no vetor de propagação pUC57.	229
Tabela 28 - Sequências dos primers utilizados para realização do protocolo de mutação sítio dirigida no sítio autoinibitório RXXD de PA0847 (R587A).....	229
Tabela 29 - Resumo de primers para clonagem de construções de PA0847 <i>P. aeruginosa</i> para vetores do duplo híbrido utilizando a técnica LIC.....	230
Tabela 30 - Primers utilizados para sequenciamento dos vetores do duplo híbrido pTRG e pBT.	230
Tabela 31 - Primers utilizados para clonagem em pETDuet das construções de HPG-1 de PA0847 e para construção HsbD-Nt.	230
Tabela 32 - Resultados de correlação (logP) maiores que 70 calculados para o alinhamento de proteínas GGDEF-EAL do banco de dados de Galperin e colaboradores.	235

Tabela 33 - Primeira parte dos dados, separados por comunidades, da variação de frequência de resíduos com e sem perturbação.....	236
Tabela 34 - Segunda parte dos dados, separados por comunidades, da variação de frequência de resíduos com e sem perturbação.....	237

LISTA DE EXPRESSÕES, ABREVIATURAS E SIGLAS

Aminoácidos	Alanina	A
	Arginina	R
	Asparagina	N
	Ácido aspártico	D
	Cisteína	C
	Ácido glutâmico	E
	Glutamina	Q
	Glicina	G
	Histidina	H
	Isoleucina	I
	Leucina	L
	Lisina	K
	Metionina	M
	Fenilalanina	F
	Prolina	P
	Serina	S
	Treonina	T
	Triptofano	W
	Tirosina	Y
	Valina	V
	Qualquer aminoácido	X
Análogo químico	molécula na qual um átomo ou grupo químico é alterado, sem que as propriedades gerais da molécula se alterem	
c-di-GMP	bis-(3',5')-di-guanosina monofosfato cíclico	
CDS	do inglês, <i><u>c</u>oding <u>s</u>equ<u>e</u>nce</i> , referente a uma sequência de DNA codificante	
DGC	usada para enzimas ou a atividade <u>d</u> iguanilato <u>c</u> iclase	
Domínio proteico	parte de uma proteína com função e enovelamento característicos, o qual em bancos de dados é definido baseado em um alinhamento múltiplo	
dRI	do inglês, <i><u>d</u>ifferential <u>r</u>efractive <u>i</u>ndex</i> , referente a variação diferencial do índice de refração	
EAL	domínio proteico ou com função de degradação do c-di-GMP para pGpG ou para ligação ao c-di-GMP. O nome vem do motivo EAL, trinca de resíduos de aminoácidos altamente recorrente/conservado no domínio EAL	
FAD	<u>f</u> lavina <u>a</u> denina <u>d</u> inucleotide	
FMN	do inglês, <i><u>f</u>lavin <u>m</u>ononucleotide</i> , ou mononucleótido de flavina	

GGDEF	domínio proteico com função de síntese do c-di-GMP ou ligação ao c-di-GMP ou GTP. O nome vem do motivo GG(D/E)EF e o nome GGDEF também pode ser usado para referenciar esse motivo
GTP	do inglês, <i>guanosine triphosphate</i> , ou guanosina trifosfato
HAMP	do inglês, <i>histidine kinases (HKs)</i> , <i>adenylyl cyclases (ACs)</i> , <i>methyl-accepting chemotaxis proteins (MCPs)</i> and some <i>phosphatases</i> , abreviação referente aos sistemas de sinalização nas quais o domínio é encontrado
HD-GYP	domínio proteico com função de degradação ou ligação ao c-di-GMP para GMP. O nome vem do motivo HD-GYP referente aos resíduos de aminoácidos HD separados da trica GYP, característicos do domínio
HK	do inglês, <i>histidine kinases</i> , referente a histidina quinases
HMM	do inglês, <i>hidden Markov model</i> , modelo estatístico que representa uma distribuição de probabilidade sobre uma sequência de observações
HPLC	técnica de cromatografia de alta performance ou alta pressão, com abreviação do inglês para <i>high-performance liquid chromatography</i>
HsbD	proteína, com nome abreviado do inglês de <i>Hpt-B-dependent secretion and biofilm</i> , por interferir na secreção e no biofilme de maneira dependente de outra proteína Hpt-B
IPTG	isopropil-b-D-galactosídeo, análogo da lactose, usado em biologia molecular para regular a transcrição da CDS de interesse
ITC	do inglês, <i>isothermal titration calorimetry</i> , referente a técnica de titulação calorimétrica isothermal
kcat	constante catalítica, igual a taxa de conversão por tempo da enzima normalizada pela quantidade da mesma
Km	constante de Michaelis-Menten, igual a concentração de substrato em que a enzima atinge metade da velocidade máxima de consumo de substrato pelo tempo
LIC	do inglês, <i>ligase independent cloning</i> , técnica de inserção de uma sequência de DNA em vetor que independe de ligase
LOV	domínio com enovelamento PAS, encontrado respondendo a luz, oxigênio e voltagem por meio dos cofatores FMN ou FAD
LS	do inglês, <i>light scattering</i> , ou espalhamento de luz
MCP	do inglês, <i>methyl-accepting chemotaxis proteins</i> , proteínas quimiotáticas dependente de metilação

mNT647	fluoróforo comercial NT-647 (NanoTemper) funcionalizado com o grupo maleimida.
Motivo	padrão encontrado recorrente/conservado em uma sequência de aminoácidos (estrutura primária). Também usado para um arranjo espacial/estrutural/tridimensional característico composto por estruturas secundárias
MST	do inglês, <i>microscale thermophoresis</i> , referente a técnica de termoforese em microescala
nhsNT647	fluoróforo comercial NT-647 (NanoTemper) funcionalizado com o grupo N-hidroxisuccinimida (NHS).
PAS	domínio ou enovelamento proteico, com abreviação derivada das proteínas nas quais foi primeiro identificado: <i>fly clock protein period</i> (<u>P</u> ER), <i>vertebrate aryl hydrocarbon nuclear translocator</i> (<u>A</u> RNT) e <i>fly developmental regulator single-minded</i> (<u>S</u> IM)
pb	pares de <u>b</u> ase, referente ao tamanho do DNA em números de nucleotídeos
PCR	do inglês, <i>polymerase chain reaction</i> , ou reação da cadeia da polimerase
PDB	do inglês, <i>protein data bank</i> , banco de dados <i>online</i> de estruturas proteicas. Também usada para o tipo arquivo de coordenadas atômicas de estruturas proteicas
PDC	enovelamento proteico, similar ao enovelamento PAS, que foi primeiro descrito nas proteínas <u>P</u> hoQ, <u>D</u> cuS e <u>C</u> itA
PDE	do inglês <i>phosphodiesterase</i> , sigla usada para a atividade fosfodiesterase ou para referenciar uma proteína com essa atividade
PFAM	do inglês, <i>protein families database</i> , banco de dados de classificação de domínios
PFstats	do inglês <i>protein families using simple statistics</i> , pacote de programas para análise de correlação entre resíduos,
pGpG	5'-fosfoguanilil-(3'-5')- guanosina
pI	ponto <u>i</u> soelétrico
PPi	pirofosfato inorgânico
q.s.p	quantidade <u>s</u> uficiente para
RMS	do inglês, <i>root mean square</i> , ou desvio quadrático médio

SDS-PAGE	do inglês <i>sodium dodecyl sulfate (SDS)–polyacrylamide gel electrophoresis</i> , referente a técnica de eletroforese em gel de poliacrilamida utilizando o detergente SDS
SEC	cromatografia de exclusão molecular, abreviação do inglês para <i>size exclusion chromatography</i>
SEC-MALS	Do inglês, <i>size exclusion chromatography with multi-angle light scattering analysis</i> , referente a técnica de cromatografia de exclusão molecular acoplada a detectores de espalhamento de luz
SUMO	do inglês <i>small ubiquitin-like modifier</i> , proteína originária de <i>Saccharomyces cerevisiae</i> e amplamente usada em biologia molecular
TIM	abreviação para <i>triose isomerase</i> , primeira proteína descrita com enovelamento barril (β/α) ₈ e, por isso, esse enovelamento também é referenciado como TIM
TM	<u>trans</u> membrana
U	unidade de enzima, ou seja, quantidade de enzima necessária para converter um micromolar de substrato em um minuto
U.A.	<u>un</u> idades <u>ar</u> bitrárias
Ulp	protease que reconhece estruturalmente e cliva SUMO, fragmento de ULP1, <i>ubiquitin-like-specific protease 1</i> , de <i>Saccharomyces cerevisiae</i>
Uniprot	do inglês, <i>Universal Protein Resource</i> , banco de dados de sequências biológicas
VC	<u>v</u> olume de <u>c</u> oluna
XDS	do inglês, <i>X-ray Detector Software</i> , programa de processamento de imagens de difração de raios X
3-AT	<u>3</u> - <u>a</u> mino-1,2,4- <u>t</u> riazol, inibidor da via sintética da histidina

SUMÁRIO

1	INTRODUÇÃO	31
1.1	Apresentação do tema em estudo	31
1.2	Motivações: relações entre sinalização bacteriana, biofilmes e o c-di-GMP	32
1.3	O que são biofilmes e por que estudá-los?	34
1.4	Microrganismos abordados no trabalho	37
1.5	Segundos mensageiros baseados em nucleotídeos	38
1.5.1	Introdução	38
1.5.2	A descoberta e as enzimas das vias do c-di-GMP	38
1.5.3	Histórico e importância de segundos mensageiros	39
1.6	Propriedades da molécula c-di-GMP	42
1.7	Como o c-di-GMP direciona mudanças em fenótipos?	43
1.8	Domínio GGDEF	46
1.8.1	Estrutura do GGDEF e a síntese do c-di-GMP	46
1.8.2	Dimerizações promovidas pelo c-di-GMP como mecanismo de inibição e captação da molécula	51
1.9	Domínio EAL	55
1.9.1	Estrutura de EAL e a hidrólise do c-di-GMP	55
1.9.2	Mecanismos de regulação de EALs conhecidos envolvendo dimerização	58
1.10	Domínios GGDEF e EAL na mesma proteína	62
1.11	Relações das vias do c-di-GMP com outras vias de sinalização	65
1.12	Histidina quinases e proteínas quimiotáticas e suas relações com o c-di-GMP	66
1.13	Regulação de domínios efetores	69
1.13.1	Mecanismos de captação de estímulos ambientais bacterianos	69
1.13.2	Estrutura e função dos domínios com enovelamento PAS	72
1.13.3	Domínios CHASE	75
1.13.4	Domínio HAMP	77
1.13.5	Transmissão da sinalização através de mudanças conformacionais por hélices	80
1.14	Sistemas estudados e objetivos do trabalho	82
1.14.1	Revisão sobre as proteínas do trabalho	83
1.14.2	Estratégias experimentais	87

2	MATERIAIS E METODOLOGIAS	89
2.1	Buscas e análises por bioinformática	89
2.2	Obtenção e manipulação dos DNAs codificantes das proteínas	90
2.2.1	Desenho de primers e plasmídeo utilizado	90
2.2.2	Protocolo padrão para PCR.....	91
2.2.3	Reação de ligação das CDS em vetor	92
2.2.4	Protocolo de transformação padrão por choque térmico	93
2.2.5	Propagação e confirmação de clones	94
2.2.6	Análise de restrição	95
2.2.7	PCR para mutação sítio-dirigida e para vetores inteiros.....	95
2.2.8	Clonagem ligase independente para vetores do duplo híbrido	96
2.3	Obtenção das proteínas	98
2.3.1	Teste de expressão e purificação.....	98
2.3.2	Protocolo de purificação e etapas de otimização	101
2.4	Produção de c-di-GMP e análises por cromatografia de fase reversa.....	103
2.4.1	Produção de c-di-GMP e purificação de c-di-GMP por cromatografia de fase reversa	103
2.4.2	Análises de c-di-GMP e pGpG por cromatografia de fase reversa.....	104
2.4.3	Cinética fosfodiesterase com EAL.....	105
2.4.4	Pré-testes de cinética com análise por fase reversa.....	106
2.5	Comprovação das biomoléculas por espectrometria de massas.....	107
2.5.1	Medidas do c-di-GMP por espectrometria de massas.....	107
2.5.2	Medidas de amostras proteicas não fragmentadas	108
2.5.3	Preparo de amostras proteicas pelo método de tripsinização.....	108
2.6	Análise de oligomerização	109
2.6.1	Análises comparativas com padrões por cromatografia de exclusão molecular (SEC) analítica	109
2.6.2	Análises por exclusão molecular acoplada a espalhamento de luz (SEC-MALS)	111
2.7	Cinética por técnica colorimétrica usando <i>kit</i> comercial	112
2.8	Medidas de interação por termoforese em microescala (MST)	115
2.9	Teste de interação com duplo híbrido bacteriano	116
2.10	Teste de interação por <i>pull-down</i>	118
2.11	Ensaio de cristalização de proteína.....	119
2.11.1	Otimização de tampão para ganho de homogeneidade proteica	119

2.11.2	Triagem por condições de cristalização.....	121
2.11.3	Expressão com selenometionina.....	122
2.11.4	Experimentos de difração de raios X.....	123
2.12	Análise em larga escala de sequências de proteínas com GGDEF e EAL.....	124
3	RESULTADOS E DISCUSSÕES	129
3.1	Estudos funcionais das diguanilato ciclasas PA0847 e HsbD.....	129
3.1.1	Planejamento experimental e produção heteróloga das proteínas do trabalho.....	129
3.1.2	Clonagem, expressão e purificação das construções de PA0847	133
3.1.3	Clonagem, expressão e purificação da construção N-terminal de HsbD	138
3.1.4	Estudos de oligomerização	140
3.1.4.1	Análise da copurificação do c-di-GMP e sua influência sobre perfis de eluição	141
3.1.4.2	Caracterizações das populações oligoméricas das construções de PA0847 por SEC analítica e SEC-MALS	145
3.1.4.3	Impacto de c-di-GMP nos equilíbrios oligoméricos das construções de PA0847	152
3.1.4.4	Impacto da força iônica da solução tampão nos equilíbrios oligoméricos das construções de PA0847.....	154
3.1.5	Estudos da atividade catalítica das construções de PA0847 contendo o domínio GGDEF	155
3.1.6	Hipóteses da regulação de PA0847 pelo domínio periplasmático (CHASE4).....	162
3.1.7	Prováveis funções estruturais do domínio PAS.....	165
3.1.8	Caracterização funcional da interação entre PA0847 e HsbD.....	168
3.3.8.1	Análises computacionais do segmento N-terminal de HsbD	168
3.3.8.2	Confirmação da interação através de ensaios de pull-down.....	170
3.3.8.3	Mapeamento de domínios de PA0847 envolvidos na interação com HsbD-Nt através de ensaios duplo híbrido.....	171
3.3.8.4	Mapeamento de domínios de PA0847 envolvidos na interação com HsbD-Nt através de ensaios de MST	172
3.1.9	Propostas de mecanismos de ativação e funcional do sistema PA0847/HsbD.....	175
3.2	Análises de coevolução indicam novos resíduos chave para GGDEF-EALs	181
3.2.1	Análises de correlação pela métrica de perturbação proposta por Bleicher e colaboradores	181
3.2.2	Análises da interface dimérica de EALs.....	191
4	CONCLUSÕES E PERSPECTIVAS	195
	REFERÊNCIAS	199
	APÊNDICES	223
	Apêndice A - Outros resultados do trabalho	223

Apêndice B - Listas de primers	227
Apêndice C - Alinhamentos completos de domínios em estudo.....	231
Apêndice D - Tabelas de dados obtidos com o PFstats.....	235
Apêndice E - Clonagem	239
Apêndice F - Espalhamento de luz acoplado a exclusão molecular (SEC MALS) ..	241
Apêndice G - Parâmetros de cinética e inibição enzimáticas	247
Apêndice H - Termoforese.....	249
Apêndice I - A termoforese em microescala no cálculo da constante de dissociação (Kd).....	251
Apêndice J - Análise de coevolução de resíduos	253

1 INTRODUÇÃO

Essa seção inicia-se com uma contextualização geral do assunto, seguida de uma descrição da relevância dos sistemas de sinalização envolvidos com o c-di-GMP, destacando o ciclo de formação do biofilme. Então, são detalhados aspectos sobre o c-di-GMP, sobre seus domínios catalíticos e sobre seus domínios regulatórios. Alguns desses dados bibliográficos serão fundamentais para formulação de hipóteses e para interpretação de resultados. No final dessa seção, é feita uma revisão bibliográfica sobre as proteínas em estudo e é apresentado os objetivos do trabalho.

1.1 Apresentação do tema em estudo

O trabalho buscou elucidar mecanismos de transdução de sinal bacterianos, que controlam os níveis intracelulares da molécula bis-(3',5')-di-guanosina monofosfato cíclico (c-di-GMP). A importância dessa molécula reside no fato de que ela regula diversos receptores, que estão envolvidos em diferentes fenótipos bacterianos. Muitos desses fenótipos estão relacionados com a alternância entre a vida livre/planctônica e a vida sésil/fixa em complexos celulares-biopoliméricos, chamados biofilmes. Os biofilmes, por sua vez, são de forte interesse industrial e medicinal.¹⁻²

Os objetos de estudo foram proteínas multidomínios e sensoras de membrana, detentoras de alguma porção receptora de sinal ambiental, associado à outra porção com domínio GGDEF e/ou EAL. Esses domínios são capazes de sintetizar e degradar, respectivamente, o c-di-GMP. Além disso, os domínios GGDEF e EAL, que não possuem resíduos para catálise (domínios degenerados), também são empregados como receptores do c-di-GMP.³ Outro destaque foi dado às proteínas nas quais foram identificados um dos domínios da família de receptores exteriores ao citoplasma denominados CHASE,⁴⁻⁷ vastamente encontrados em vias de sinalização bacterianas.

Dentre outras proteínas abordadas no presente estudo (apêndice A), resultados promissores foram obtidos com as de *Pseudomonas aeruginosa*, codificadas nos genes PA0847 e PA3343, sendo que PA3343 já é descrita como HsbD.⁸ Uma das inovações do estudo é que PA0847 tem estrutura e função desconhecidas. Além disso, PA0847 e HsbD potencialmente compõem um sistema do tipo específico de direcionamento da sinalização por c-di-GMP, por vezes referenciado como *local pool*, que vem sendo cada vez mais comum em casos da

literatura.⁹⁻¹² A interação entre PA0847 e HsbD foi encontrada por experimentos de triagem de interação por duplo híbrido, realizado pelo nosso grupo de pesquisa.¹³

Adicionalmente, o presente trabalho abordou um aspecto pouco compreendido nas vias de sinalização mediadas por c-di-GMP, sobre a regulação e atuação proteínas contendo ambos domínios GGDEF-EAL em tandem, que possuem alta incidência em genomas bacterianos e que comumente estão associados a membrana. Análises de bioinformática visaram a busca e a compreensão de resíduos de aminoácidos altamente frequentes (conservados) e/ou correlacionados (coevolúidos). Esses resíduos poderiam estar envolvidos na regulação de domínios GGDEF e EAL na mesma proteína, muitas delas transmembranares.

Para esse propósito, foram feitas análises de alinhamentos de diversos exemplares de GGDEF e EAL em tandem. Os resíduos encontrados poderiam, por exemplo, dar indícios da existência de algum mecanismo de robustez^a dessas proteínas, tópico que nunca foi abordado no contexto das vias do c-di-GMP, apesar de ser um aspecto comum de enzimas bifuncionais.¹⁵ Outro fato, que poderia vir a ser elucidado, através da análise de proteínas GGDEF-EAL, é sobre os mecanismos regulatórios alternativos para dimerização de EALs, sendo que dímeros são conhecidos por serem a forma ativa. Uma das bases de comparação sequencial e estrutural foi a STM3615 de *Salmonella enterica*, que teve a porção citoplasmática caracterizada estruturalmente e *in vitro* em outro trabalho inédito, desenvolvido no nosso grupo de pesquisa.¹⁶

1.2 Motivações: relações entre sinalização bacteriana, biofilmes e o c-di-GMP

Bactérias são extremamente hábeis em responder adequadamente a estímulos ambientais, tais como temperatura, presença de nutrientes, substâncias agressivas e outros.^{2,17-18} Entender os mecanismos bacterianos de sensoriamento e resposta pode vir a ser útil em diversos aspectos. Esses sistemas podem vir a compor dispositivos biosensores ou servir como modelo na elaboração dos mesmos.¹⁹⁻²¹ E isso com a vantagem da menor complexidade dos sistemas sensores bacterianos, em comparação aos presentes em eucariotos.

Outras aplicações de sistemas de sinalização envolvem o aperfeiçoamento ou a fundamentação de metodologias de controle bacteriano. Ou seja, técnicas que bloqueiem ou mimetizem sinais ambientais captados por bactérias de maneira que essas se comportem a favor

^a Robustez aqui é referente a habilidade de uma rede bioquímica funcionar corretamente, mesmo frente a flutuações das espécies químicas da rede.¹⁴

do ser humano. Algo similar ao que as plantas fazem naturalmente²² e que apenas recentemente vem sendo compreendidas e exploradas.¹⁷

Figura 1 - Esquemática das etapas do ciclo do biofilme bacteriano e da localização das proteínas em estudo.

O ciclo de formação do biofilme pode ser dividido em cinco etapas.²³⁻²⁴ Ele se inicia com os microrganismos livres, que se aderem em uma superfície de maneira reversível (i). Em uma segunda etapa (ii), as células passam a se fixar na superfície de maneira irreversível e a formar agregados celulares. Em seguida, inicia-se a produção da matriz biopolimérica e a formação de microcolônias (iii). Então ocorre o amadurecimento do complexo, com a diversificação das células, coordenada por sinais do *quorum sensing*^b, definição de microambientes e até a formação de canais hídricos (v). O biofilme maduro pode abranger uma diversidade de espécies, incluindo diferentes reinos da vida, fenômeno conhecido como coagregação.²⁶ A liberação de células, natural ou com o desprendimento de fragmentos do biofilme, permite a disseminação do(s) microrganismo(s) no ambiente (v).²⁷⁻²⁸ Em destaque, o tema em estudo: proteínas transdutoras de sinal multidomínios que respondem a estímulos ambientais, alterando a homeostase do c-di-GMP, importante controlador de fenótipos bacterianos.

Fonte: Elaborada pelo autor.

^b *Quorum sensing* é a sinalização entre bactérias dependente da densidade celular. A alta densidade celular permite o acúmulo de moléculas chamadas autoindutoras que se difundem de uma célula para outra. Os autoindutores em gram-negativas são N-acil-homoserina lactonas (AHL) enquanto que gram-positivas são encontrados peptídeos cíclicos com papel de autoindutores.²⁵

Um dos intermediadores mais comuns entre sinais ambientais e efetores dos fenótipos associados a virulência e ao ciclo de formação do biofilme é a molécula bis-(3',5')-di-guanosina monofosfato cíclico, abreviada como c-di-GMP, di-GMP cíclico ou apenas como cdG. De maneira geral, altos níveis intracelulares de c-di-GMP coordenam a maquinaria celular bacteriana para o biofilme enquanto que, em células planctônicas, são encontrados baixos níveis do dinucleotídeo cíclico (Figura 1).^{24,29}

Alguns detalhes moleculares de como o c-di-GMP promove mudanças no metabolismo bacteriano já são bem descritos na literatura e serão abordados a seguir. Outro fator atrativo da pesquisa vem da ausência de c-di-GMP na vasta maioria dos eucariotos. Isso faz das vias de sinalização mediadas por c-di-GMP ótimos alvos para o desenvolvimento de pequenas moléculas, capazes de interferir de maneira específica no metabolismo bacteriano.²⁹⁻³⁰

Dentre os fenótipos guiados por estímulos ambientais se destacam aqueles associados a virulência³¹ e ao ciclo entre a formação microbiana em comunidade, denominada biofilme, e a vida livre unicelular, também chamada planctônica. Esses fenótipos chamam a atenção não apenas pelo número, mas também pelos significativos impactos, positivos e negativos, no cotidiano humano.

1.3 O que são biofilmes e por que estudá-los?

Biofilmes são compostos por células de uma ou mais espécies de bactérias, fungos e algas.³²⁻³³ Grande parte do biofilme é formada pela matriz, na qual se encontram embebidos os microrganismos. Essa formação ocorre aderida à uma superfície, que pode ser de origem biológica ou não. A composição da matriz do biofilme engloba água, lipídeos e diversos biopolímeros extracelulares, ou EPS, abreviação do inglês para *extracellular polymeric substances*.^c Esses biopolímeros incluem proteínas, DNA extracelular (DNAe) e vários tipos de exopolissacarídeos.² A proporção de substâncias e o tipo de biopolímeros que compõe a matriz variam com a composição celular do mesmo e depende do ambiente no qual o biofilme se desenvolve.^{32,34} Microrganismos em diferentes localizações em um biofilme apresentam diferenças na expressão de moléculas de superfície, em fatores de resistência, nas taxas de utilização de nutriente e na produção de fatores de virulência.³⁵

^c É possível encontrar também a abreviação EPS para exopolissacarídeos.

Polissacarídeos da matriz são encontrados com a função de reserva de nutrientes, retenção hídrica, barreira que retarda ou impede a penetração de agentes antimicrobianos e manutenção da estrutura. DNAs podem ser vistos formando redes ou filamentos no biofilme, com funções estruturais diversas, tais como a de adesão celular e coesão do biofilme. A proximidade das células no biofilme facilita trocas gênicas, com destaque para aqueles de resistência. Proteínas são responsáveis por metabolizar o EPS para absorção de nutrientes e liberação de células. Outros tipos de proteínas e glicoproteínas que se ligam a carboidratos (lectinas) conectam as bactérias a matriz, além de evitar que enzimas deixem o biofilme.

Células isoladas apresentam maior produção de fatores de virulência enquanto biofilmes conferem as células constituintes um maior grau de resistência a eliminação.³⁶⁻³⁷ O aumento da resistência bacteriana está relacionada a uma fase semelhante à de dormência das células no biofilme. Também ocorre o retardo na penetração de agentes antimicrobianos pela matriz, conferindo tempo para expressão de genes de resistência. Além disso, a matriz possui enzimas que degradam antibióticos. (Figura 2A e Figura 2B).

Figura 2 - Exemplificações de danos causados por biofilmes.

A) Superfícies inertes, como as de dispositivos clínicos e a de tecidos mortos são candidatos proeminentes a desenvolver biofilmes. Bactérias na forma livre são mais suscetíveis ao ataque de fagócitos e a ação de antibióticos; B) O mesmo não ocorre com células do biofilme. O acúmulo de enzimas fagocitárias acaba danificando o tecido saudável em torno da infecção. A liberação de bactérias acaba disseminando a infecção; C) Um dos grandes problemas na indústria é a formação de biofilmes em tubulações, fonte de corrosão (figura), eliminação e entupimento. Companhias gastam bilhões anuais combatendo tais problemas.

Fonte: Adaptada de COSTERTON; STEWART; GREENBERG³⁸; COSTERTON; STEWART²⁷.

Na indústria, por exemplo, biofilmes são uma recorrente causa da contaminação persistente na manipulação de alimentos⁴⁰ e da corrosão e entupimento de tubulações (Figura

2C).³⁹ Ademais, existe um grande interesse tecnológico em relação a aplicações dos biofilmes. Pode-se citar o uso em biosensores,⁴¹ em células voltaicas MFCs, (*microbial fuel cells*)⁴² e em estratégias de biorremediação.^{20,43} Isso agrega interesse em conhecer os mecanismos de controle do ciclo de formação e dispersão do biofilme.

INFECÇÕES RELACIONADAS A DISPOSITIVOS MÉDICOS

Figura 3 - Esquematização dos dispositivos médicos susceptíveis ao crescimento de biofilme e patologias agravadas e/ou causadas por essa formação microbiana.

Fonte: Adaptada de LEBEAUX et al.⁴⁴

Os biofilmes na medicina são, entre outros, responsáveis por agravar infecções ao se estabelecer em superfícies de tecidos como a dental, pulmonar (agravada em pessoas com fibrose cística) e em válvulas cardíacas. O biofilme também é responsável pelo deterioramento de dispositivos médicos como cateteres, próteses e lentes⁴⁵ (Figura 3). Metade dos casos de infecções nosocomiais^d nos Estados Unidos estão associados com dispositivos médicos permanentes⁴⁶ (Figura 3). Outros dados norte-americanos ainda mostram que, do total de

^d Expressão usada para infecções adquiridas em hospitais ou clínicas.

infecções humanas, 65% estão relacionadas com biofilmes, de acordo com os CDCs (*Centers for Disease Control and Prevention*)⁴⁷ ou 80%, de acordo com NIHs (*National Institutes of Health*).^{44,48}

1.4 Microrganismos abordados no trabalho

Dada a relevância dos temas de sinalização bacteriana e dos fenótipos que controlam alguns sistemas de sinalização de diferentes bactérias, foram escolhidos para o presente estudo proteínas provenientes de patógenos importantes. Um deles é *P. aeruginosa*, um importante patógeno oportunista de humanos imunocomprometidos e agente de graves infecções pulmonares de pacientes com fibrose cística⁴⁹ (Figura 4A). Esse microrganismo é conhecido por usar o biofilme como estratégia para aumentar a resistência e, para coordenar a infecção, utiliza vastamente vias do c-di-GMP. Dada sua importância, o microrganismo comumente é usado como modelo para estudos.

Figura 4 - Microrganismos dos quais as proteínas foram abordadas no presente estudo.
Fontes: Adaptada de ORGANS...⁵⁰; SCIENCE...⁵¹; ATKINS⁵²; BERGAMIN.⁵³

Em 2017 a organização mundial de saúde (OMS) cita a cepa de *P. aeruginosa*, resistente aos beta-lactâmicos carbapenens, como prioridade crítica para alvo de desenvolvimento de fármacos.⁵⁴ Além disso, diferentes espécies do gênero *Pseudomonas* são conhecidas pela sua versatilidade em colonizar diferentes ambientes e infectar diferentes organismos^{24,55-56} o que demonstra a variedade e complexidade dos sistemas de sinalização desse microrganismo. Outro material de estudo vem de *Salmonella enterica*, microrganismo que recebe atenção por infectar humanos e animais. Ele é agente de problemas muito comuns de saúde pública mundial, como a gastroenterite e complicações infecções mais severas, como bacteremia e a febre entérica ou tifoide⁵⁷ (Figura 4B).

1.5 Segundos mensageiros baseados em nucleotídeos

1.5.1 Introdução

O c-di-GMP atua em diferentes fenótipos, especialmente aqueles vinculados com o biofilme. E o c-di-GMP faz parte de um grupo de moléculas denominadas segundas mensageiras. De forma geral, moléculas sinalizadoras dessa classe tem sua concentração intracelular alterada em resposta a presença ou ausência, no meio extracelular, de determinados fatores, chamados primeiros mensageiros. Segundos mensageiros são úteis para que o sinal consiga ser transmitido e/ou amplificado para dentro da célula. Eles se ligam em diferentes alvos específicos promovendo mudanças estruturais nos mesmos, o que resulta na alternância no estado funcional. Dentre os diversos mecanismos de sinalização celular, segundos mensageiros compõem cascatas de sinalização essenciais, tanto em procariotos quanto em eucariotos.⁵⁸⁻⁵⁹

No geral, para cada rota de sinalização baseada em segundos mensageiros, existe um domínio proteico específico empregado para a sua síntese e outro para degradação. Os primeiros mensageiros são responsáveis por controlar, de maneira alostérica, a atividade desses domínios de biossíntese e metabolismo de segundos mensageiros. Essa regulação geralmente ocorre por meio da atuação de domínios proteicos extras dessas enzimas. Em particular, para vias baseadas em nucleotídeos mono ou dicíclicos, grupo que engloba o c-di-GMP, a síntese é realizada pelas chamadas ciclases. A degradação, por sua vez, é feita enzimas que pertencem a um extenso grupo de esterases, as fosfodiesterases^e (PDEs).

1.5.2 A descoberta e as enzimas das vias do c-di-GMP

A primeira identificação do c-di-GMP veio apenas em 1987,⁶¹ como regulador alostérico da celulose sintetase, hoje conhecida como BscA. Sua descoberta foi fruto inesperado de trabalhos de Benziman e colaboradores durante a década de 1980, que buscavam compostos determinantes para atividade da celulose sintase de *Gluconacetobacter xylinus*, na época denominada *Acetobacter xylinum*. Nas vias do c-di-GMP, a síntese é realizada por diguanilato ciclases (DGCs), que são caracterizadas por possuírem o domínio GGDEF, que deve possuir

^e Apesar do mesmo nome e atuarem na mesma função química, fosfodiesterases compõe diferentes famílias⁵⁸ e são designadas para atuar sobre diferentes moléculas. Eventualmente, de maneira inespecífica, podem hidrolisar outras moléculas.⁶⁰

todos os resíduos catalíticos presentes. A partir de duas moléculas de GTP, é formado o c-di-GMP e duas moléculas de pirofosfato (PPi) como subproduto. Fosfodiesterases específicas do c-di-GMP, por sua vez, são caracterizadas por possuírem ou o domínio EAL ou o domínio HD-GYP. Esses domínios catalisam, respectivamente, a hidrólise do c-di-GMP para 5'-fosfoguanilil-(3'-5')- guanosina (pGpG) e guanosina monofosfato (GMP). Para isso, esses domínios utilizam como cofator um cátion divalente, sendo o mais comum reportado o Mg^{2+} . Detalhes estruturais e da catálise de GGDEFs e EALs serão abordados com detalhes nas próximas seções (1.8 e 1.9).

1.5.3 Histórico e importância de segundos mensageiros

Diferentes classes de moléculas são usadas como segundas mensageiras, tais como Ca^{2+} , fosfatidilinositois, alguns gases (NO, CO e H_2S) e nucleotídeos cíclicos. Dentre os nucleotídeos cíclicos que agem como segundos mensageiros (Figura 5), o primeiro a ser isolado⁶² e identificado⁶³ foi o 3',5'-monofosfato cíclico de adenosina (cAMP). Ele foi apenas no final da década 50, através de estudos com extratos de hepatócitos, produzido em resposta à presença de hormônios epinefrina/adrenalina e glucagon. Logo a seguir, em 1963, foi publicada a identificação do cAMP em *Escherichia coli*⁶⁴ e *Bacterium liquefaciens*.⁶⁵ Outro segundo sinalizador, o 3', 5'-monofosfato de guanosina, ou cGMP, como o cAMP, primeiramente descrito em eucarioto, foi detectado na urina de ratos em um trabalho de 1963,⁶⁶ para posteriormente ser detectado em bactérias, em 1974.⁶⁷⁻⁶⁸

A produção de cAMP, em eucariotos, é controlada pelos clássicos receptores acoplados a proteína G (GPCRs - *G-protein-coupled receptors*). Esses receptores também são denominados receptores de sete-hélices (7TM - *seven-transmembrane receptors*). Receptores 7TM se distinguem principalmente pela porção receptora extracelular, cada uma específica para hormônios, neurotransmissores, íons, odorantes, luz e outros.⁶⁹ De maneira simplificada, esses receptores desencadeiam a ativação de uma série de enzimas, dentre elas as adenilatos ciclases (AC), que convertem o ATP em cAMP.

O cAMP possui uma variedade de receptores, dentre eles os mais comuns são outras quinases A (PKA, *protein kinases A*), capazes de fosforilar^f e assim desencadear a alteração do estado de uma variedade de proteínas. GPCRs também são encontrados controlando a conversão de fosfoinositol 4,5-bifosfato (PIP₂) para diacilglicerol (DAG) e inositol trifosfato

^f A fosforilação, ou quinase, no contexto biológico, consiste na transferência de um grupo fosfato geralmente fornecido por um ATP para a cadeia lateral de resíduo de uma proteína alvo.

(IP3), outros segundos mensageiros essenciais em eucariotos. Tanto DAG quanto IP3 estão envolvidos, por exemplo, na regulação do Ca^{2+} , outro elemento essencial em uma variedade de rotas de sinalização.⁷⁰ Em bactérias, o cAMP é amplamente reconhecido por participar da preferência catabólica por glicose em *E. coli*. Somente na última década, o cAMP vem ganhando destaque como regulador de diversas outras vias bacterianas.

Enquanto que, em eucariotos, a sinalização mediada por moléculas baseadas em nucleotídeos é predominantemente baseada em mono-nucleotídeos cíclicos cAMP e cGMP, bactérias usam um repertório muito maior desses sinalizadores, incluindo (p)ppGpp (análogos não-cíclicos de GTP e GMP) e os di-nucleotídeos cíclicos c-di-GMP e c-di-AMP (Figura 5). Usa-se (p)ppGpp para se referir a duas moléculas, guanosina 5'-trifosfato,3'-difosfato (pppGpp) e a guanosina 5'-difosfato,3'-difosfato (ppGpp). Foram primeiramente reportados como pontos mágicos (*magic spots*), em 1969.⁷¹ Isso porque foram encontrados de maneira inesperada em análises de moléculas marcadas com fosfato radioativo (^{32}P) de extratos de *E. coli*,²⁹ sob escassez de aminoácidos. Inclusive, por serem frequentemente produzidos em resposta a condições de estresse, (p)ppGpp são conhecidos também como alormônios. A produção dessas moléculas promove uma resposta estrigente em bactérias, alteração transcricional que prolonga a fase estacionária de divisão celular, ao inibir a replicação de DNA e síntese de lipídeos entre outros, para priorizar biossíntese de aminoácidos e vias proteolíticas para obtenção de energia.⁷²

Apenas em 2008, após a descoberta do c-di-GMP em 1987,⁶¹ foi reportado pela primeira vez outro segundo mensageiro, o bis-(3'-5')-di-adenosina monofosfato cíclico, molécula abreviada para c-di-AMP ou di-AMP cíclico. Foi encontrado por acaso, como uma densidade eletrônica extra na resolução da estrutura de DisA. Essa proteína realiza o escaneamento da integridade do DNA (*DNA integrity scanning protein*) de *Thermotoga marítima*, que possui um homólogo em *Bacillus subtilis*. E isso foi possível porque o c-di-AMP copurifica com DisA.

A proteína DisA possui o domínio com mesmo nome, que caracteriza uma diadenilato ciclase (DAC). Ela se organiza em octâmeros, que produzem c-di-AMP enquanto fazem a varredura no DNA. Quando um DNA ramificado é encontrado, provável produto de dano, a produção de c-di-AMP cessa. Dessa forma é desencadeado o atraso da esporulação em *B. subtilis*.⁷³ Ao longo dos anos foi descoberto ainda que o c-di-AMP atua em diversas vias bacterianas, como por exemplo na homeostase de fosfato, osmoproteção, expressão gênica, vias de resistência a antibiótico e até na formação de biofilme.¹ Ao contrário do c-di-GMP, o c-di-

AMP é mais recorrente em bactérias gram-positivas. Enquanto isso, o (p)ppGpp participa de múltiplos processos, tanto em gram-positivas quanto em gram-negativas.

Figura 5 - Estruturas moleculares de nucleotídeos usados encontrados como segundos mensageiros clássicos.

Fonte: Cedida pelo doutorando Anacleto Souza.

A descoberta de outro segundo mensageiro veio em 2012, com a enzima DncV (VC0179), capaz de sintetizar c-di-GMP, c-di-AMP e preferencialmente o inédito híbrido 3',5' GMP-AMP cíclico, ou 3'-3' cGAMP (encontrado também como cGMP-AMP ou como cGAMP). A DncV foi descoberta ao acaso, quando se investigava ilhas genômicas importantes para patogenicidade de *Vibrio cholerae*.⁷⁴

Enquanto que, até o momento, (p)ppGpp, c-di-AMP e 3'-3' cGAMP foram identificados exclusivamente em procariotos, c-di-GMP faz parte as vias de sinalização de pelo menos um organismo eucarioto, o *Dictyostelium discoideum*.^g Logo após, em 2013, foram publicados as descobertas, em eucarioto, do cGAMP⁷⁵ e da cGAMP sintase (cGAS), ativada por DNA exógeno no citosol.⁷⁶ A única diferença do cGAMP de eucarioto para o de procarioto é que a ligação do carbono 3' da desoxirribose ligada adenina ocorre para do carbono 2' da

^g *D. discoideum* é unicelular e entra em um ciclo pluricelular sob condições de estresse ambiental, ciclo iniciado com a coordenação da sinalização por cAMP extracelular.⁵⁹

desoxirribose ligada a guanina⁷⁷, por isso a abreviação 2'-3' cGAMP (Figura 5). Experimentos sugerem que o receptor de cGAMP em eucariotos, a STING (do inglês, *stimulator of interferon genes*)^h, também capaz de responder também a c-di-GMP e c-di-AMP.⁷⁹⁻⁸⁰

1.6 Propriedades da molécula c-di-GMP

A concentração do c-di-GMP intracelular pode variar dependendo da bactéria analisada e das condições de cultivo. Medidas experimentais apontam algo em torno de 0,1 a 2,0 pmol/(mg de célula),⁸¹ o que equivale a uma concentração intracelular bacteriana de 10 μ M na célula.⁸² O c-di-GMP em solução é estável, não sofrendo hidrólise ou degradação espontânea. Possui massa molecular de 690,09 Da e coeficiente de extinção molar (ϵ) de 26 100 ($M^{-1}cm^{-1}$, no 260 nm)⁸³ ou de 23 700 ($M^{-1}cm^{-1}$, no 253 nm).⁸⁴⁻⁸⁵ Ele se apresenta em equilíbrio em solução entre formas monoméricas, diméricas, tetraméricas e até octaméricas (Figura 6).

Figura 6 - Arquitetura e equilíbrio dos estados oligoméricos apresentado em solução do c-di-GMP.

Molécula do dímero de c-di-GMP extraído da estrutura de PleD (PDB 2v09).

Fonte: Adaptada de WANG et al.⁸⁶

Foi visto que esse equilíbrio é sensível a concentração de c-di-GMP e a presença de cátions.⁸⁶⁻⁸⁷ Formas do c-di-GMP monoméricas, diméricas e mais recentemente tetraméricas já foram observadas ligadas a estruturas proteicas.⁸⁸⁻⁸⁹ Espectros de ressonância magnética nuclear (RMN) indicam grandes agregados de c-di-GMP a altas concentrações, em torno de 30 mM.⁹⁶ Tanto domínios GGDEF assim como o EAL podem utilizar como cofator o íon Mg^{2+} , mais comumente citado na literatura, cátion com concentração celular bacteriana de aproximadamente 2 mM.⁹⁰

^h STING também é encontrada como MITA, ERIS, MPYS or TMEM173.⁷⁸

1.7 Como o c-di-GMP direciona mudanças em fenótipos?

A regulação das atividades catalíticas antagônicas de DGCs e PDEs modulam os níveis intracelulares do sinalizador c-di-GMP que, por sua vez, é reconhecido por uma grande variedade de receptores, levando à progressão de diferentes processos fisiológicos (Figura 7).

Figura 7 - Esquematização de exemplos das vias de sinalização do c-di-GMP.

A síntese do c-di-GMP é feita através de DGCs, a partir de dois GTPs. A degradação leva a formação de dois GMPs ou em pGpG (c-di-GMP linearizado) quando a catálise é feita pelas fosfodiesterases com EAL e HD-GYP, respectivamente. Os domínios GGDEF e EAL sem atividade catalítica, ou degenerados, mostrados na figura como GGDEF^{deg} e EAL^{deg}, foram encontrados como domínios receptores de c-di-GMP. Na esquematização das enzimas, a esfera marcada com X representa os diversos domínios extras associados com ativação e desativação das atividades DGC e PDE. Abaixo estão representados os diversos exemplos de receptores (destacados em círculos), o que a molécula induz (em preto) e a resposta fenotípica (em vermelho).

Fonte: Adaptada de HENGGE.¹⁰

Desde sua descoberta, vários alvos alostéricos do c-di-GMP vem sendo amplamente descritos e em diversas vias de sinalização bacteriana. O c-di-GMP se liga a seus receptores em diferentes configurações e oligomerizações e promove nos receptores mudanças conformacionais e de estados de oligomerização.⁸⁸⁻⁸⁹ Domínios proteicos específicos podem ser encontrados empregados como receptores da molécula. O primeiro exemplo descoberto foi o do domínio PilZ,⁹¹ presente em BcsA, proteína do trabalho pioneiro do c-di-GMP,⁶¹ e que é encontrado em diversos outros exemplos.

Domínios GGDEF, EAL e HD-GYP, que não são capazes de sintetizar ou degradar o c-di-GMP, são amplamente encontrados como receptores da molécula. Nesses casos, os domínios são conotados como degenerados. Receptores de c-di-GMP estão constantemente sendo descobertos, como por exemplo o domínio GIL.⁹² O c-di-GMP atua em outras vias de exopolissacarídeos, além da via da celulose, como na produção de alginato, tendo como alvo a proteína Alg44. O c-di-GMP atua, entre outros, em alvos como fatores de transcrição que, por exemplo, atuam na virulência e produção de flagelo (Clp e FleQ), na divisão celular (PopA), na adesão celular (LapD), na locomoção (YcgR) e outros. O c-di-GMP regula ainda diretamente os níveis de tradução através de sítios de mRNA, característicos por reconhecer uma variedade de moléculas, localizados na extremidade 5' do biopolímeros (Figura 7).

É comum em um único genoma bacteriano encontrar codificado múltiplas proteínas contendo GGDEFs, EALs assim como domínios receptores do c-di-GMP. A maneira como o c-di-GMP alcança os seus alvos é um tema em aberto. O primeiro modelo tem princípio intuitivo e é citado como *global pool*. O c-di-GMP, produzido por qualquer DGC, pode livremente se difundir no citoplasma e indiscriminadamente se ligar a qualquer um de seus receptores ou a qualquer uma de suas fosfodiesterases, com EAL ou HD-GYP (Figura 8A). Entretanto, muitos estudos vêm mostrando que o impacto na concentração de c-di-GMP, promovido por diferentes DGCs, pode ser de diferentes maneiras, num mesmo fenótipo. Isso corrobora para um segundo modelo, referenciado em alguns trabalhos como *local pool*.⁹⁻¹⁰ Nesse modelo, uma ou mais enzimas são responsáveis por direcionar especificamente o c-di-GMP para um receptor ou um grupo deles¹¹ (Figura 8A).

As explicações encontradas na literatura de como se alcança essa especificidade é localizando os metabolizadores e receptores, produzindo-os de maneira simultânea, ou seja, a nível transcricional ou ainda pelas diferentes afinidades por c-di-GMP pelos receptores.⁹ Uma terceira forma é controlando o destino do c-di-GMP através de complexos proteicos. Apesar de ainda não bem elucidado, o *local pool* é a ideia mais defendida, pois explica algumas aparentes contradições experimentais.

A)

B)

Figura 8 - Mecanismos propostos para o direcionamento do c-di-GMP para receptores e experimento que corrobora para o modelo específico (*local pool*).

A) Ilustração comparativa entre os modelos de entrega do c-di-GMP para os receptores. No modelo não específico, ou *global pool*, o c-di-GMP alcança os alvos (setas) por difusão simples no citoplasma. No direcionamento específico, ou *local pool*, os mecanismos de entrega não são totalmente compreendidos. Podem ser controlados por exemplo, por complexos proteicos; B) O exemplo da figura traz o resultado de um trabalho de larga escala de *V. cholerae*. O c-di-GMP produzido por diferentes DGC impactam com diferentes intensidades na formação do biofilme. Isso fica evidente quando é traçada a correlação das diferentes DGC isoladamente (r^2 mais próximo de 1).

Fonte: A) Elaborada pelo autor; B) de MASSIE et al.¹¹

Um dos exemplos que podem ser explicados pelo direcionamento específico são experimentos que mostram a mesma quantidade de c-di-GMP produzido por diferentes DGCs impactando com diferentes intensidades no mesmo fenótipo (Figura 8B).⁹³ Outro exemplo vem com duas proteínas, RoeA e SadC, de *P. aeruginosa*. Tanto a cepa sem RoeA, como gene mutado ($\Delta roeA$), quanto outra incapaz de expressar SadC ($\Delta sadC$), são observados níveis reduzidos de c-di-GMP, se comparada a cepa original com essas duas proteínas (ou selvagem). A mutação $\Delta roeA$ promove alto decréscimo de exopolissacarídeo Pel e nenhuma alteração na produção de flagelo. Enquanto isso, nas células com $\Delta sadC$ não ocorre variação do Pel, mas a função do flagelo é alterada. Outro exemplo de controle do impacto do c-di-GMP, através de complexos, vem com a LapD e GcbC, de *Pseudomonas fluorescens*. GcbC possui um domínio

GGDEF ativo e uma porção receptora do periplasma. Já LapD utiliza um domínio GGDEF apenas como receptor da molécula. Através de mutações simples e mutações compensatórias nas hélice $\alpha 4$ do GGDEF de GcbC ou na hélice $\alpha 2'$ do EAL de LapD, foi possível demonstrar que essas proteínas interagem, por intermédio dessas estruturas secundárias.¹²

1.8 Domínio GGDEF

1.8.1 Estrutura do GGDEF e a síntese do c-di-GMP

O nome do domínio GGDEF, responsável por formar o c-di-GMP, vem do motivo de resíduos com a sequência GG(D/E)EF, motivo proteico encontrado abreviado apenas como GGDEF (Figura 10A). Os resíduos do motivo possuem altos índices de conservação e compõem a maior parte daqueles empregados na catálise. Mecanismos atômicos sobre a catálise e autorregulação de DGCs foram introduzidos com a proteína PleD de *Caulobacter vibrioides* (*Caulobacter crescentus*).⁹⁴⁻⁹⁵ Por isso, esse texto, como em outros, utiliza como referência a numeração de aminoácidos dessa proteína, com exemplares de estruturas no PDB 1w25 e 2v0n.

O motivo GGDEF, que compõe o sítio ativo de uma DGC, é comumente abreviado para sítio A, ou sítio ativo. Entretanto, outros resíduos, além daqueles da assinatura GGDEF, são necessários para catálise, como será visto a seguir. Mesmo assim, em um único GGDEF, está presente apenas metade de um sítio ativo na prática. Isso porque, na síntese, são necessárias duas moléculas de guanosina trifosfato (GTP), sendo que cada domínio GGDEF é responsável por coordenar um único GTP. Portanto, são necessários dois domínios GGDEF para formação de um sítio ativo para atividade DGC. Outra restrição essencial para síntese de c-di-GMP é a presença coordenação do GTP interagindo no sítio A do GGDEF através de cátions divalentes, sendo o mais reportado o Mg^{2+} . Na reação, ocorre a liberação de dois pirofosfatos (PPi) como subproduto. Além disso, os GGDEFs devem estar simetricamente justapostos na catálise por um eixo ordem 2 e os resíduos catalíticos devem estar corretamente posicionados.^{3,96}

Os quatro primeiros resíduos do motivo GGDEF formam um motivo estrutural do tipo grampo β , do inglês, β -*hairpin*, localizada entre as fitas centrais $\beta 2$ e $\beta 3$ (Figura 9). O grampo do GGDEF possui morfologia incomum pois, dentre outras razões, não apresenta uma ligação hidrogênio na cadeia lateral entre o nitrogênio do resíduo da posição $i+3$ com o oxigênio de i ($NH_{i+3} \rightarrow O_i$), típica de grampos e voltas β . Além disso, os oxigênios da cadeia lateral em i , $i-1$ e $i-2$ se encontram perpendiculares ao plano formado por $\beta 2$ e $\beta 3$, resultando em uma torção no final de $\beta 2$ no sentido de $\beta 3$. Isso é atribuído a presença de uma arginina (R366), comum na

posição i-2, a qual os nitrogênios da cadeia lateral mantêm ligações com as cadeias principais de duas glicinas, G368 e G369, das posições i e i+1 (Figura 9).

Figura 9 - Estrutura característica do motivo GGDEF.

A estrutura do grupo formado pelos resíduos GGDEF é peculiar em relação aos grupos, estrutura que é mantida por contatos com uma arginina (R), anterior na sequência. Na figura ainda é mostrado o motivo RXXD, motivo regulatório, tratado na seção 1.8.2. Apesar de próximo sequencialmente, esse motivo se localiza em uma face oposta do enovelamento em relação ao GGDEF.

Fonte: Adaptada de SCHIRMER.⁹⁷

Trabalhos pioneiros com DGCs mostraram que todos os cinco resíduos do motivo GGDEF são necessários para atividade. Entretanto, posteriormente foram relatadas certas flexibilidades por trocas na posição G368. Por exemplo, VCA0965 de *V. cholerae* possui sítio A como AGDEF⁹⁸ e foi relatada com atividade. Também ECA3270, de *Pectobacterium atrosepticum*, que possui a sequência SGDEF como sítio A, é capaz de sintetizar c-di-GMP. Essa mesma DGC permanece ativa trocando-se o primeiro resíduo do motivo para G ou A.⁹⁹ Não é bem esclarecida a função do resíduo F371, mas foi visto que trocar esse resíduo por G, L, I, V, R, D, E, N, S ou C elimina a atividade de WspR.¹⁰⁰ Em outro exemplo, a substituição F371A em HmsT de *Yersinia pestis* leva a instabilidade proteica.¹⁰¹ Apesar disso, foi demonstrado *in vivo* e *in vitro* a atividade de DgcP, uma das três DGCs de *B. subtilis* e que possui GGEEL como motivo catalítico. Em WspR, ainda foi proposto que a posição 367 deve ser uma Y, caso o sítio A tiver os resíduos GGEEF.¹⁰⁰

Domínios GGDEFs possuem por volta de 170 resíduos¹⁰² e apresentam um enovelamento sanduiche alfa-beta ($\beta\alpha\alpha\beta\beta\alpha\beta$), com a folha antiparalela central de topologia 2-3-1-4-5, referente a ordem das fitas beta na folha (Figura 10B). Essa folha é cercada geralmente por cinco hélices. A reação em dois passos em DGCs, cada passo referente a um dos ataques nucleofílicos da ribose, foi evidenciada logo pelo artigo pioneiro de Benziman e

colaboradores.⁶¹ Esse identificou o composto de transição 5'-pppGpG e comprovou a possibilidade em usá-lo como substrato. Posteriormente esse modelo foi refinado pelos trabalhos estruturais com as enzimas PleD, WspR, entre outras. Com WspR, por exemplo, foi relatado o uso como substrato o 5'-pppGpG.¹⁰³ Um modelo similar de reação de dois passos foi proposto para formação do c-di-AMP.¹⁰⁴ Os fosfatos β e γ do GTP⁹⁴, que dão origem ao grupo lábil PPI, são acomodados pelo loop P, entre $\beta 1$ e $\alpha 1$, e pelas duas primeiras glicinas da assinatura do GGDEF (G368 e 369) da unidade vizinha.⁹⁷ Esses fosfatos estão coordenados por K442, R446 e um Mg^{2+} , referenciado como Mg^{2+} (B).^{95,i} Esse íon, por sua vez, é coordenado pelo resíduo D327 e o quarto resíduo motivo GGDEF, E370.

Para ocorrer a ciclização, dois domínios carregados com GTP e cofatores devem formar um dímero transiente compatível com o prosseguimento da reação. Uma vez posicionados corretamente, o oxigênio da hidroxila 3'-OH da ribose, de um dos GTPs, é desprotonada, possivelmente por uma água.⁹⁷ O ânion formado, então, realiza um ataque nucleofílico sobre o fosfato α do segundo GTP. Na face oposta do GTP no qual ocorre o ataque, se encontra o fim da cadeia lateral do terceiro resíduo do sítio A, que são os resíduos D327 e E370, do motivo GGDEF. Em PleD, esse resíduo corresponde a E371. Esse resíduo é apontado como o catalítico. Espacialmente próximo ao E371, se localiza um segundo íon Mg^{2+} , ou Mg^{2+} (A), que atua como uma base geral na reação e que também forma interação com D327. O ataque resulta na liberação de um pirofosfato (PPI). Simetricamente, a ribose do segundo GTP deve atacar o fosfato α do primeiro GTP para fechar o ciclo de fosfatos do c-di-GMP.^{95,97}

Na reação, K332 é usada para estabilizar cargas do estado de transição pentafosforil e do grupo lábil PPI,⁹⁵ portanto, bem conservada em DGCs. Os resíduos N335 e D344 se ligam e conferem especificidade a base guanil do GTP. Segundo dados cristalográficos, as bases guanina não mudam de posição durante os passos da reação. Os resíduos N335 e D344 se encontram nas hélices $\alpha 1$ e $\alpha 2$, conjunto de hélices que já foi chamada na literatura de torre (Figura 10C).⁹⁷

A primeira estrutura do dímero funcional veio apenas com a resolução da estrutura de DgcZ⁹⁶, também conhecida como YdeH⁸⁵ (Figura 12). Essa proteína de *E. coli* possui um domínio sensor de Zn^{2+} , que inativa o GGDEF na presença do ligante. Estão depositados no PDB o domínio GGDEF de DgcZ em complexo com o produto c-di-GMP no sítio A (PDB: 3tvk) e a estrutura inteira inibida com Zn^{2+} e em complexo com GTP α S (análogo químico do GTP, com hidrólise muito lenta). A ordem de grandeza para a constante catalítica (k_{cat})

ⁱ A nomeação do magnésio tem analogia a descrição proposta para adenilatos ciclases.

$\text{NH}_{i+3} \rightarrow \text{O}_i$, apresenta uma ligação $\text{NH}_{i+4} \rightarrow \text{O}_i$ o que torna a volta maior, por isso, do inglês, *wide-turn* ou WT.⁹⁷ Logo a seguir se encontra a asparagina N299 que ainda faz contatos importantes entre a hélice α_0 do GGDEF com a hélice posterior, que conecta o GGDEF a um domínio anterior na cadeia polipeptídica (Figura 11). Pode se especular que essa volta, somada com N299 e outro resíduo bem conservado, o R300, poderiam estar envolvidos na estabilização do dímero cataliticamente competente. Entretanto, a função desses elementos não é elucidada até o momento.

Figura 11 - Representação das estruturas dos motivos GGDEF e DXLT, característicos de domínios GGDEF.
Fonte: Adaptada de SCHIRMER.⁹⁷

O domínio GGDEF apresenta diversas similaridades estruturais e funcionais com nucleotídeo ciclases do tipo III, grupo que abrange adenilatos ciclases (AC), sintetizador do cAMP. Similaridades entre AC e DGCs são utilizadas inclusive para justificar relações evolutivas.¹⁰⁵ No geral essas enzimas possuem um núcleo estrutural comum, que possui correspondências (Figura 10B). Em adenilatos ciclases (ACs), diguanilato ciclases (DCs) e DNA polimerases possuem um grampo análogo ao sítio A de DGCs. Em DCs o motivo característico do sítio ativo é o XGDXF/Y, que permite mais variações de resíduos que o motivo GGDEF. ACs e DCs utilizam, como DGCs, o loop P para coordenar os fosfatos β e γ do substrato. O motivo DXLT é encontrado em outras estruturas como a AC como DDLS (PDB 2ak3) e STK como DKLT (PDB: 3uc3).⁹⁷ A presença do motivo DXLT em outros sistemas de sinalização reforça a importância da sua função que ainda não elucidada.

É possível encontrar o domínio GGDEF degenerado usado para ligar GTP, como observado para proteína FimX de *P. aeruginosa* e PdeA de *C. crescentus*, também citada como CC3396. Nesses casos, o ligante promove a ativação das proteínas.¹⁰⁶⁻¹⁰⁷ O motivo autoinibitório RXXD do GGDEF, tema da próxima seção, pode ser encontrado como receptor

do c-di-GMP em domínios GGDEF degenerados, como é o caso de PelD, de *P. aeruginosa*.¹⁰⁸ Um EAL degenerado também pode ter a função de ligar c-di-GMP, como aparece na mesma FimX.¹⁰⁹

Figura 12 - Dímero do domínio GGDEF na conformação ativa exemplificada pelo sensor de zinco DgcZ. Uma das unidades está destacada em verde claro e outra em verde escuro. A estrutura da direita é a visão superior da imagem esquerda. O eixo de simetria é destacado com uma seta. A localização do domínio ligador de zinco CZB (*chemoreceptor zinc-binding domain*) está esquematizada em cinza na figura da esquerda.

Fonte: Elaborada pelo autor.

1.8.2 Dimerizações promovidas pelo c-di-GMP como mecanismo de inibição e captação da molécula

Outro motivo característico de domínios GGDEFs é o R³⁵⁹XXD³⁶², localizado em uma volta que antecede o GGDEF, entre $\alpha 2$ e $\beta 2$. Apesar da proximidade sequencial, na estrutura esse motivo está antipodal ao motivo GGDEF. O RXXD é referenciado como sítio inibitório (sítio I) ou sítio inibitório primário (sítio Ip).¹¹⁰ O segundo nome vem de seu papel essencial em dois modos de ligação ao c-di-GMP observados, não apenas na inibição não competitiva do produto (Figura 13), mas também em formas de ligação do c-di-GMP em receptores (Figura 14). Esses modos de ligação são referenciados aqui como modo de inibição 1 e 2. Ambos os modos de ligação foram observados em vários exemplos e na mesma proteína. Em comum, nos dois modos o c-di-GMP se liga na forma de dímero, representado como (c-di-GMP)₂.

O sítio Ip em PelD compreende, além dos resíduos R359 e D362, o resíduo R390. No primeiro modo de inibição, o sítio Ip interage com (c-di-GMP)₂ que, por sua vez, também interage com R313, proveniente de um segundo domínio GGDEF (Figura 13 A). Esse resíduo R313 compõe o chamado sítio inibitório secundário, que aqui será referenciado como sítio Is(1). Essas interações permitem a formação de um homodímero de GGDEFs. Nesse dímero inativo os sítios estão em faces opostas e cataliticamente impedidos de sintetizar c-di-GMP. O aspartato D362 nesse modo de interação está provavelmente envolvida na especificidade pelo ligante,

Figura 14 - Modos de dimerização em DGCs e receptores promovidos pela ligação do c-di-GMP dimérico.

Na figura, R1 e R2 representam os sítios Ip (do RXXD) e Is respectivamente. Os traços em verde representam as interações entre as argininas do sítio e os grupos aromáticos do grupo guanosina.

Fonte: Adaptada de GIARDINA et al.¹¹¹

YfiN, por exemplo, de *P. fluorescence*, também conhecida como TpbB e de código PA1120, é capaz de converter GTP em c-di-GMP, mas é incapaz de se ligar ao c-di-GMP, portanto, não ocorre autoinibição na enzima. Ela é um transdutor de membrana com receptor externo e contendo na parte citossólica um domínio HAMP seguido de um GGDEF (PDB: 4iob). A ausência da interação com c-di-GMP foi atribuída a ausência do sítio Is(1). A atividade regulatória nesse caso foi atribuída exclusivamente ao domínio receptor periplasmático, mesmo possuindo o sítio inibitório RXXD.¹¹¹ No segundo modo de inibição também ocorre a participação do sítio Ip. Ele interage com (c-di-GMP)₂ que, dessa vez, interage ainda com resíduos de outro domínio da mesma cadeia polipeptídica. No caso de PleD esse segundo grupo de resíduos são provenientes do domínio REC e são R148 e R178. Essas posições compõem um segundo tipo de sítio secundário, abreviado nesse texto como Is(2) (Figura 13 B).

PleD possui na porção N-terminal dois domínios REC. Experimentos mostraram que sua forma ativa é adquirida com a fosforilação do primeiro REC, que pode ser mimetizada *in vitro* com o uso de BeF₃⁻. A estrutura disponível no PDB ativa é encontrada no modo autoinibido, ou seja, no modo de inibição 2. Uma vez que a região conectora entre o segundo REC é flexível, foi postulado um equilíbrio entre dímero de GGDEFs na forma ativa e forma

inibida.⁹⁵ O modo de inibição 2 é aquele utilizado também para o reconhecimento da molécula. Nesse modo de ligação heterodimérica, outros domínios fazem o papel similar ao do domínio REC em PleD (Figura 14).¹¹¹

Sistemas de regulação em GGDEFs pelo produto são importantes pois impedem que a bactéria consuma GTP de maneira indiscriminada. Mesmo em XCC3486 de *Xanthomonas campestris* que não possui o RXXD foi demonstrada a inibição competitiva pelo produto ($K_i = 6 \mu\text{M}$). Mesmo em DgcZ o sítio I se apresenta como RXXE, mas ainda é capaz de ligar ao c-di-GMP ($K_i = 4 \mu\text{M}$). Nesse caso, outro fator para a inibição ganha importância, que é a mudança estrutural exercida pelo domínio extra com sítio para Zn^{2+} , ou domínio CZB (de *chemoreceptor zinc-binding domain*). Segundo o modelo proposto, a ligação do zinco desencadeia em DgcZ uma separação do dímero de GGDEFs, o que distancia resíduos do sítio ativo e, assim, impedindo a catálise.

A ordem de grandeza para K_d do c-di-GMP pelo sítio inibitório é da ordem de 1 a $10^{-1} \mu\text{M}$.⁸⁸ DGCs como PleD e DgcA^{46,56} apresentam constante de inibição em ordem de grandeza parecida, de $\sim 1 \mu\text{M}$. Devido a essa afinidade por c-di-GMP proteínas produzidas em sistemas de expressão heterólogos clássicos são copurificados com a molécula ligada. Algo semelhante inclusive permitiu a primeira observação do c-di-AMP. Um exemplo o qual ocorre isso é o da proteína tDGC (TM1788),¹¹² do extremófilo *T. marítima*, que possui apenas o domínio GGDEF, sem domínio regulatório. Tal fato também foi observado para WspR,¹⁰³ em *P. aeruginosa* (PA3702), que possui um domínio REC seguido de um GGDEF. Em ambas os casos, para se obter proteínas apo foram aplicados protocolos de incubação com a fosfodiesterase RocR de *P. aeruginosa* (PA3947), proteína com domínio REC seguido de um domínio EAL. tDGC é purificada na forma de dímeros. Após ser submetida a um tratamento com fosfodiesterase, se apresenta predominantemente monomérica. A mutação em tDGC no sítio Ip R158A também resulta em proteínas monoméricas.¹¹² WspR é copurificada com c-di-GMP e se apresenta forma dimérica nas formas ligada e apo. Entretanto, a forma ligada é um dímero alongado e a forma apo é um dímero compacto, formas associadas com a regulação de WspR.¹⁰³ Em um trabalho posterior, foi mostrado que a atividade dessa proteína é dependente da formação de tetrâmeros.¹¹³

1.9 Domínio EAL

1.9.1 Estrutura de EAL e a hidrólise do c-di-GMP

De maneira similar ao GGDEF, o nome do domínio EALs vem dos aminoácidos com alto grau de identidade entre exemplares do mesmo. Também são necessários ainda muitos outros resíduos para catálise (Figura 15A) além da tríade de resíduos EAL. Recentemente, foi descoberto que oligoribonucleases Orn são capazes de hidrolisar pGpG em GMP, fechando o ciclo de degradação do c-di-GMP.^{3,114} Como o trabalho tem como foco apenas o domínio EALs, o domínio HD-GYP não será descrito.

As primeiras estruturas EAL foram resolvidas na forma apo e são de YkuI de *B. subtilis* (PDB:2bas) depositada em 2005¹⁰⁶ e TBD1265, referenciada também como TdEAL⁹⁰, de *Thiobacillus denitrificans* (PDB:2r6o) datada de 2007.⁸² O domínio EAL contém aproximadamente 240 resíduos e possui enovelamento clássico da topoisomerase TIM, ou $(\beta/\alpha)_8$ (Figura 15B e Figura 15C). Essa estrutura é composta por 8 fitas centrais paralelas que formam um barril e essa estrutura é cercada por oito hélices. Como ocorre em barris TIM, o EAL possui o sítio catalítico localizado na face C-terminal do barril, região que abrange as extremidades C-terminais das fitas que compõe o enovelamento. TBD1265, por ser uma das primeiras estruturas de uma EAL ativa e resolvida com o substrato (PDB: 3n3t), será usada como referência nesse trabalho. Entretanto, a numeração hélices e fitas segue padrão para barris TIM, muito comum em outros trabalhos.^{90,115-116} Dentre as diferenças do barril do EALs para o barril clássico, temos que as duas primeiras fitas $\beta 1$ e $\beta 2$ possuem aproximadamente o dobro do tamanho das demais fitas, sendo que a $\beta 1$ é antiparalela as outras fitas do barril (Figura 15C).

Baseados nas estruturas cristalográficas, proporam-se dois modelos para a hidrólise. O primeiro é baseado em um único cátion divalente.⁸³ O segundo modelo, mais aceito, mostrou que na catálise são usados dois desses íons, algo clássico para fosfodiesterases.^{82,115} A maior eficiência é obtida com Mn^{2+} , sendo que podem ser utilizados Co^{2+} ou Mg^{2+} . Entretanto é conhecido que Ca^{2+} e Zn^{2+} inibem o domínio EAL. No mecanismo clássico para fosfodiesterases um dos metais auxilia na ativação de uma água, ou seja, na remoção de um H^+ , formando assim o nucleófilo hidroxila (OH^-). Enquanto isso, o segundo metal estabiliza o estado de transição. Contudo, no mecanismo proposto de dois metais para o domínio EAL os dois íons cooperam nas duas tarefas.

O trabalho com TBD1265 identificou 10 resíduos com alta frequência em EALs. Quando algum desses resíduos é trocado por alanina ocorre a diminuição drástica da atividade

fosfodiesterase. Entretanto, esse mesmo trabalho mostrou que apenas 7 dessas posições mutadas praticamente extinguem a atividade PDE. Em RocR 7 resíduos de aminoácidos foram identificadas como essenciais, ou seja, a mutação por alanina diminui extremamente a constante catalítica k_{cat} (decréscimo $< 10^5$ vezes em relação a proteína sem mutações).^{83,117} Essas posições serão abordadas durante a descrição do domínio (tabela 1).

Tabela 1 - Resumo de resíduos do domínio EAL importantes para a atividade.

TBD1265	E523	R527	E546	N584	E616	-	D646	D647	K667	E703	Q723
RocR	E175	-	-	N233	E265	E268	D295	-	K316	E352	-

*Em vermelho estão os resíduos que quando trocados eliminaram a atividade da respectiva enzima e em preto estão os resíduos que quando mutados fornecem uma queda significativa. A posição E268 de RocR corresponde a posição E619 de TBD1265.

Fonte: Elaborada pelo autor.

Usando a numeração da TBD1265, o primeiro cátion divalente (Me1) é coordenado pelos resíduos E523, N584, E616 e D646 além da água catalítica (A1) e um dos oxigênios do c-di-GMP (O3'). Esses elementos formam um octaedro ao redor do íon Me1 (Figura 17).¹¹⁵ O oxigênio O3' é aquele ligado ao fosfato P1 do c-di-GMP. Cada um dos resíduos que coordenam Me1 vem de uma fita diferente do barril e os quatro compõem o grupo dos sete resíduos essenciais para a atividade. O segundo íon (Me2) interage também com a água A1, um dos oxigênios do fosfato P1, com E703, D646 e D647. Em estruturas como RocR de *P. aeruginosa*, e YahA de *E. coli*, foi observada a coordenação do Me2 também pelo correspondente D669 de TBD1265. Os magnésios teriam função de coordenar a água, diminuir seu pKa para gerar o íon hidróxido, polarizar a ligação entre P1 e o oxigênio ligado ao carbono 3' (O3') e estabilizar o estado de transição.⁸³

Na literatura, E703 é mais bem aceito como resíduo catalítico. Isso por ser um dos resíduos catalíticos essenciais e por ter um dos seus oxigênios da cadeia lateral alinhado com A1, P1 e O3', observação feita em RocR.⁸³ Observação análoga foi feita em BlrP1 de *Klebsiella pneumoniae*, na qual E703 coordena A1 apenas na conformação ativa, posição mimetizada por pH alto em experimentos de cristalografia.¹¹⁵ Na catálise, a hidroxila formada a partir da água A1 realiza um ataque nucleofílico sobre o fosfato P1. Ocorre então a quebra da ligação com O3', grupo lábil da reação. Propôs-se que esse oxigênio então recebe um próton de uma segunda água (A2) presente no sítio ativo. Essa doação final do próton é feita por um par histidina-glutamato em outras fosfodiesterases, par de resíduos ausentes no domínio EAL (Figura 15D).⁸²

Figura 15 - Distribuição dos elementos do EAL sequencial e tridimensional.

A) Distribuição sequencial de estruturas secundárias e resíduos chave do domínio EAL. Em destaque estão resíduos relevantes para função do domínio; B) Distribuição aproximada dos componentes de estrutura secundária, destacando a topologia das folhas que compõe o barril TIM. As hélices em vermelho, assim como o motivo DDFG(T/A)GYSS são elementos conhecidos por participam da dimerização entre EALs; C) Estrutura tridimensional do domínio EAL, representado pela estrutura do dímero TBD1265; E) Alguns dos resíduos importantes são destacados em vermelho estão mostradas as cadeias laterais dos mesmos, destacadas em roxo.

Fonte: Elaborada pelo autor.

Na assinatura EAL estão os resíduos E523, que coordena Me1, e o L525, usado como suporte para o anel de fosfato do c-di-GMP na reação. Próximo sequencialmente está R527, que interage com o segundo fosfato (P2) do c-di-GMP. Tanto o motivo EAL quanto o L525 se localizam na $\beta 2$, o que mostra a importância dessa fita. Em torno da junção entre as hélices $\alpha 2$ e $\alpha 2'$, existe um bolsão que acomoda um dos grupos guanina(G1) do c-di-GMP. Em TBD1265, os resíduos do bolsão correspondem a V526, I542, I551, L554 e V558, resíduos que ainda podem variar em outros EALs. Essa região ainda abrange o resíduo E546, visto como essencial para a atividade. Alguns resíduos conferem especificidade pelo grupo guanina(G2) do c-di-GMP, com destaque para o mais recorrente E706.⁸²

Outro resíduo essencial é a K667, encontrada orientando A1 e realizando contatos com E523, E616 e Q723 formando uma rede de interação (EEKQ),⁸² importante na ativação da água catalítica, no posicionamento⁸³ e estabilização dos estados desprotonados desses resíduos,⁸² que coordenam Me1. No trabalho que avaliou a atividade de TBD1265, resíduos conservados como Q723, localizada no final da fita $\beta 8$, T618, T650 e R527 quando mutados diminuem, mas não anulam, a atividade fosfodiesterase. Os resíduos T618 e T650 não possui função descrita.

O trabalho com TBD1265 apontou D647 como ativador de A1, pois observaram que a mutação em D647 TBD1265 diminui drasticamente a atividade. O trabalho conclui também que esse fato explicaria porquê YkuI é inativa, apesar de possuir todos os resíduos essenciais para atividade, exceto D647.⁸² Como contraponto, em RocR a mutação em D647 apenas diminui o k_{cat} em 33 vezes. A importância do resíduo está relacionada com motivo que ele constitui D⁶⁴⁶DFG(T/A)GYSS. É proposto que esse motivo participe da regulação de EALs⁶⁰ por isso, esse motivo será referenciado aqui como motivo regulatório.

Os dois primeiros resíduos do motivo regulatório são D646 e D647 empregados para coordenar Me1 e Me2 e que se localizam no fim da fita $\beta 5$. Os demais resíduos formam o loop 5, entre $\beta 5$ e $\alpha 5$. O resíduo essencial E619 (que corresponde a E265 de RocR), localizado no loop 4, entre $\beta 5$ e $\alpha 5$. Esse resíduo está relacionado coordenação do loop 5, importante na dimerização, tema da próxima seção.

1.9.2 Mecanismos de regulação de EALs conhecidos envolvendo dimerização

Três fatores adicionais, além da presença dos resíduos catalíticos, são cruciais para atividade de EALs e que parecem estar vinculados. Um deles é a conservação de E619, o segundo é conservação do aqui referenciado motivo regulatório D⁶⁴⁶DFG(T/A)GYSS e o

terceiro é a formação de oligômeros, com dímeros como na TBD1265,⁸² ou tetrâmeros, como ocorre no dímero de dímeros de RocR.¹¹⁸

Exemplos mostram como a conservação de E619, estabilizador do loop 5, pode interferir na atividade. Trocando o resíduo em RocR ocorre a formação de estados alta ordem de oligomerização (HMWO- *high molecular weight oligomer*) e queda ou anulação da atividade, reforçando a participação do resíduo na oligomerização (tabela 2).⁶⁰ Comparando estruturas obtidas sob diferentes pHs, no qual um pH baixo mimetiza a forma ativa da proteína, é possível notar mudanças estruturais mais significativas no E619, além da hélice $\alpha 4$. Observações semelhantes foram feitas com outra proteína com EAL, PA2567.¹¹⁵

Tabela 2 - Dados de exemplos do efeito de mutações em resíduos importantes sobre parâmetros cinéticos.

Proteína (resíduos originais do loop 5)	Local da Modificação	Resíduo alterado (Equivalente em TBD1265)	k_{cat} (s^{-1})	K_m (μM)	K_i (μM)	Oligomerização
RocR - motivo D ²⁹⁵ DFGAGYSS		Sem modificação (selvagem)	0,67±0,03	3,2± 0,3	-	Tetrâmero
	Estabiliza loop 5	E268A(E619A)	<10 ⁵	-	-	HMWO
		E268Q (E619Q)	(1,5±0,1) x10 ⁻³	0,3± 0,1	-	HMWO e tetrâmero
	Loop 5	D296A(D647A)	(2,1±0,3) x10 ⁻²	8,6±2,8	-	Tetrâmero
		F297(F648A)	0,02±0,01	0,7±0,2	-	HMWO
		S302(S653A)	0,14±0,07	5,5±2,6	10,9±3,0	Tetrâmero
PA2567 - motivo D ⁴⁹⁶ DFGTGYSN		Sem modificação (selvagem)	0,39±0,03	5,2±1,3	-	Dímero
	Loop 5	E464A (E619A)	1,1±0,5	13± 5	30±10	Dímero

*HMWO: *high molecular weight oligomer*.

Fonte: Adaptada de RAO et al.⁶⁰

Poucos são os exemplos de EAL degenerados que possuem o motivo regulatório conservado, o que sugere sua relação com a atividade.⁶⁰ Um dos exemplos que demonstram a importância da conservação do motivo regulatório é dado com AxDGC2 de *Acetobacter xylinum*. Essa proteína é originalmente dimérica e capaz de hidrolisar outros substratos inespecíficos, mas incapaz de hidrolisar o c-di-GMP. Trocando o motivo regulatório não

conservado de AxDGC de DNFGKGITVL para DDFGTGYTVL, é possível obter uma AxDGC2 capaz de degradar c-di-GMP.⁶⁰

Figura 16 - Visão superior do dímero de EALs, com destaque para o eixo de simetria (TBD1265, PDB:3n3t).

A) Dímero de EALs, no qual cada unidade está colorida por uma cor (laranja e marrom); B) Modelo simplificado do dímero funcional. Cada um dos domínios é representado por um cilindro e as faces C-terminais e N-terminais dos barris estão destacados (C_B e N_B respectivamente). A interface de dimerização está marcada com uma estrela com a letra D; C) Dímero de EALs com uma visão superior do domínio, destacando os nomes das hélices e a interface dimérica. Na figura são destacados em rosa escuro a cadeia lateral dos resíduos importantes para dimerização Y e S, pertencentes ao motivo regulatório.

Fonte: Elaborada pelo autor.

Tanto a proteína inteira, que é dimérica, quanto o domínio isolado EAL (YkuI^{EAL}), que é monomérico¹⁰⁶ são inativas *in vitro*. Nas estruturas de YkuI apo (PDB: 2bas) ou ligada (PDB: 2w27) o resíduo não coordena o loop degenerado. Uma mutação em YahA no resíduo D647A, do motivo regulatório, promove inativação da enzima, mesma posição não conservada naturalmente em YkuI. RocR com mutações nas posições equivalentes D647 e S653 alteram parâmetros cinéticos da proteína sem alterar a forma tetramérica. Isso reforça a importância da conservação do loop para a atividade. Em outro exemplo, RbdA¹¹⁹ de *P. aeruginosa* possui apenas alguns dos resíduos do motivo regulatório presentes, apresentando a sequência DDFCAGMS. Entretanto, a dimerização nessa proteína ocorre entre $\alpha 6'$ e a hélice sinalizadora que conecta os outros domínios citoplasmáticos PAS e GGDEF em RbdA.

Por fim, vários exemplos mostram a necessidade da formação de dímeros para a atividade, com interface formada geralmente por resíduos do loop 5, que abrange o motivo DDFG(T/A)GYSS, e das hélices $\alpha 5$ e $\alpha 6'$. A conformação dimérica ativa canônica foi observada não apenas no domínio EAL de referência TBD1265⁸² (Figura 16A) e de RocR,⁸³ MorA,¹¹⁶ BlrP1,¹¹⁵ YahA,⁹⁰ YkuI¹⁰⁶ e LapD (EAL receptor de c-di-GMP).¹²⁰ No dímero, as unidades estão justapostas e com os sítios catalíticos, na interface C do barril TIM, apontados para direções opostas (Figura 16B).

Estudos estruturais com MorA, de *P. aeruginosa*, forneceram um modelo de como a dimerização induzida pelo ligante interfere na atividade. Nessas estruturas, a MorA, utiliza a sequência YS do motivo regulatório para dimerização, assim como visto em outras estruturas como em BlrP1 e TBD1265. Por outro lado, nas estruturas monoméricas de MorA foi observada uma hélice 5 mais longa, que engloba os resíduos YS, que no dímero é desestruturada e faz contatos com os resíduos YS da unidade EAL vizinha (Figura 16C). Na estrutura do monômero os demais resíduos do motivo regulatório ficam fora do sítio catalítico, incluindo D646 e D647. A extensão da hélice no monômero impõe ainda choques estéricos que impedem a formação do homodímero de EAL. Pelo papel na regulação, $\alpha 5$ foi nomeada como hélice repressora.¹¹⁶

Em YahA, uma mutação em uma serina da hélice $\alpha 6'$ para um triptofano diminui drasticamente a dimerização e também a atividade.⁹⁰ Entretanto essa mutação não compromete a afinidade de YahA por c-di-GMP. O domínio EAL de YahA isolado (YahA^{EAL}) é capaz de hidrolisar c-di-GMP e se apresenta em equilíbrio monômero-dímero. O c-di-GMP desloca esse equilíbrio para a forma dimérica e cátions, de maneira curiosa, deslocam o equilíbrio para monômeros. Na estrutura, é possível observar que E619 coordena o loop com a presença do c-di-GMP. Cátions deslocam a proteína para o estado monomérico e na estrutura não há o posicionamento de E619 para o loop⁹⁰ (Figura 17).

Figura 17 - Estruturas cristalográficas de YahA, com e sem o ligante, e BlrP1 com o ligante.

O resíduo E619, que corresponde a E235 em YahA e E275 em BlrP1, resíduos destacados com tracejado vermelho. O c-di-GMP é capaz de deslocar o equilíbrio de YahA para dímeros enquanto a adição de apenas cátions desloca o equilíbrio para monômeros. Na estrutura com c-di-GMP, o E619 coordena o loop 5 (destacado em roxo) como ocorre em estruturas ativas, por exemplo, em BlrP1.

Fonte: Adaptada de SUNDRIYAL et al.⁹⁰

1.10 Domínios GGDEF e EAL na mesma proteína

É comum encontrar o domínio GGDEF seguido, em tandem, por um EAL ou mesmo HD-GYP.¹²¹ Aproximadamente 1/3 de todas as proteínas com GGDEFs também possuem um EAL e 2/3 de todas as proteínas nas quais aparece um EAL também é encontrado um GGDEF (observações próprias usando o Uniprot e material de referência¹²¹). A maioria dos exemplos abordados neste texto são de proteínas com essa arquitetura (Figura 18).

Para outras enzimas, que são bifuncionais e com atividades opostas, a bifuncionalidade é responsável pela robustez do controle via que a mesma participa.¹⁴ Diversos modelos já foram propostos para sistemas robustos, alguns deles já foram divididos de acordo com a dependência ou não da formação de uma estrutura quaternária e se as catálises ocorrem num sítio comum ou em sítios separados.¹⁵ Entretanto, para proteínas com GGDEF e EAL conservados não existem trabalhos voltados para investigação da validade ou proposta de um modelo robusto.

Outra importância de enzimas detentoras tanto do GGDEF quanto do EAL vem de trabalhos que investigaram vias de especificidade de regulação envolvendo domínios da homeostase ou de ligação do c-di-GMP. Nos testes de interação entre domínios dos genomas de *E. coli*¹²² e de *P. fluorescens*⁹ foram constatados que proteínas com GGDEF e EAL formam

hubs nas redes de interações constatadas. Isso mostra a importância dessas proteínas para as vias específicas de c-di-GMP.

Figura 18 - Resumo de arquiteturas de domínios de algumas das proteínas citadas nesse texto, grande parte detentoras tanto do domínio GGDEF quanto do EAL.

Domínios GGDEF e EAL marcados com faixas escuras indicam domínios degenerados. Domínios GGDEF marcados com apenas uma única faixa escura indica são aqueles que não possui o RXXD. Domínios EAL marcados com uma faixa escura única são aqueles que não possuem o motivo regulatório D⁶⁴⁶DFG(T/A)GYSS.

Fonte: Elaborada pelo autor.

É possível encontrar em bancos de sequências proteínas com GGDEF e EAL preditos catalíticos, apenas um dos domínios ativos ou outras que ambos domínios são inativos. FimX de *P. aeruginosa* e LapD *P. fluorescens* são exemplos em que ambos, GGDEF e EAL, são degenerados. Esses domínios são empregados como receptores de c-di-GMP ou GTP. LapD responde a c-di-GMP, utilizando para isso um EAL degenerado. A ligação ao c-di-GMP nessa proteína promove um rearranjo estrutural dos domínios GGDEF e EAL que é transmitida para a porção periplasmática. Essa porção então captura a protease LapG, que atua sobre uma adesina LapA. Dessa forma é possibilitada a atuação de LapA na adesão celular (Figura 19).¹²⁰

MorA e RbdA são exemplos já descritos de proteínas tanto com a atividade DGC quanto PDE. Foram constatadas as atividades em solução tanto DGC e PDE da porção de MorA,¹¹⁶ que engloba os domínios GGDEF e EAL (resíduos 978–1409), e da fração citoplasmática de RbdA¹¹⁹ (resíduos 232–818). Inclusive a atividade PDE de RbdA é estimulada na presença de um análogo de GTP, substrato do domínio GGDEF.

Figura 19 - Esquemática do mecanismo de atuação da LapD.

O c-di-GMP se liga ao domínio EAL da proteína e promove uma mudança estrutural drástica na conformação dos domínios. Essa mudança é transmitida para o domínio periplasmático, que recruta a proteína LapG, uma protease de LapA.¹²⁰

Fonte: Adaptada de NAVARRO et al.¹²⁰

Um dos trabalhos do nosso grupo de pesquisa resultou na estrutura de um GGDEF degenerado seguido de um EAL conservado, provinda da proteína STM3615, resíduos 242-668.¹⁶ A disposição espacial dos domínios GGDEF e EAL se assemelha a encontrada em MorA. Os domínios GGDEF nessas estruturas estão em uma conformação próxima da observada no modo de inibição 1 (Figura 20).

Figura 20 - Estruturas dos dímeros de GGDEF-EAL de STM3615 (resolvida pelo nosso grupo) e MorA.¹¹⁶

STM3615 é um exemplo de GGDEF degenerado junto com um EAL conservado, enquanto MorA apresenta tanto o GGDEF quanto o EAL ativos.

Fonte: Elaborada pelo autor.

Uma das diferenças das estruturas de STM3615 e MorA com outras GGDEF-EAL com os dois domínios degenerados, como LapD e FimX, é o maior tamanho entre as hélices conectoras entre os domínios GGDEF e EAL. Uma particularidade da construção de STM3615 que possui apenas o EAL, resíduos 392-668, é a forma dimérica em solução. Para outros EALs isolados, como YahA ou MorA, é observada a presença de monômeros em solução que, com e sem um equilíbrio para dímeros, dependente da presença de c-di-GMP ou cátions como Mg^{2+} , ocorre a observação de dímeros. Esses dados levaram a concluir que a hélice conectora é a responsável por alternar estados em proteínas com arquiteturas GGDEF-EAL com atividade DGC, PDE ou ambas.¹⁶

1.11 Relações das vias do c-di-GMP com outras vias de sinalização

Vias de sinalização que usam o c-di-GMP possuem várias relações e compartilham vários elementos, não apenas com vias baseadas em outros segundos mensageiros, como as do cAMP e com vias de fosforilação. Vários desses elementos em comum estão bem descritos apenas para sistemas fosforilativos, também chamados quinase. Além disso, sistemas fosforilativos e de vias de segundos mensageiros podem compor uma única rota de sinalização. Inclusive, o número de proteínas sensoras em geral de um microrganismo já foi relacionado a capacidade adaptativa do mesmo, através de um definido “Q.I. bacteriano”.¹²³ Por isso, se faz necessário descrever um pouco sobre esses outros sistemas, pois podem auxiliar a entender os mecanismos de ativação e desativação de DGCs e PDEs do c-di-GMP.

As vias mais comuns de fosforilação são compostas por proteínas histidina quinases (HKs, *histidine kinases*), quimiotáticas aceptoras de metil, encontradas também como quimiorreceptoras (MCPs, *methyl accepting chemotaxis protein*) e serina/treonina quinases, ou STKs, *serine threonine kinases*.¹²⁴ Uma das semelhanças de sistemas de sinalização bacteriano é o compartilhamento de vários domínios acessórios regulatórios.¹²⁵ Os mais comuns são os domínios REC, HAMP, PAS, GAF, CHASE, MASE, BLUF, Hélice-S e PilZ (Figura 21A). E isso ocorre com uma extrema variedade de combinações e num mesmo genoma (Figura 21B). O fenômeno da variedade de composições de domínios regulatórios e catalíticos (ou efetores) é denominado embaralhamento de domínios, ou do inglês, *domain shuffling*.¹²⁶

Figura 21 - Ilustração da variedade de possíveis combinações de domínios de vias de sinalização bacterianas (A) e algumas das proteínas das vias do c-di-GMP de *P. aeruginosa* (B), exemplificando essa variedade.

A) A composição de domínios para sequências proteicas que se pode encontrar depositadas em bancos *online* é enorme. Os domínios regulatórios ilustrados na figura são os mais frequentes. Domínios GGDEF e EAL não catalíticos (marcados com faixas) podem aparecer como a função regulatória; B) Algumas das proteínas contendo GGDEF e EAL identificadas no genoma de *P. aeruginosa*. O número total de proteínas que possuem GGDEF, EAL, HD-GYP ou PilZ é próximo de 40, dependendo da cepa analisada de *P. aeruginosa*. Na figura estão marcadas com * as proteínas que possuem o motivo autoinibitório (inibido pelo próprio produto) denominado RXXD. Faixas azuis verticais representam porções preditas transmembrana. Pequenos destaques em rosa, na ponta esquerda da representação de cada proteína, indicam peptídeos sinais. Variações do assinaturas sequenciais (motivos) que caracterizam os domínios GGDEF e EAL são mostradas nos blocos que representam os respectivos domínios. Na ausência da atividade catalítica, os domínios GGDEF ou EAL potencialmente servem como receptor de c-di-GMP.

Fonte: A) Elaborada pelo autor; B) Adaptada de CHRISTEN et al.¹²⁷

1.12 Histidina quinases e proteínas quimiotáticas e suas relações com o c-di-GMP

HK são caracterizadas por dois domínios. Um deles possui atividade fosfatase e é encontrado abreviado como CA, por ser o catalítico, ligando-se a ATP (*catalytic/ATP-binding*). Também é encontrado nos bancos de dados SMART e Pfam como HATPase_c, mas é referenciado como domínio GHKL (Abreviação para *Gyrase, Hsp90, Histidine Kinase, MutL*). O segundo domínio se localiza anterior ao domínio GHKL e é chamado transferidor de fosfato (DHP, abreviação para *donor histidine phosphotransfer*), encontrado também como HisKA no

Pfam ou SMART. O conjunto GHKL e DHp também é referenciado na literatura como KCM, abreviação para *kinase control module*.¹²⁸

HKs alterna entre atividade quinase e fosfatase. Na atividade quinase em HK ocorre a transferência de um grupo fosfato de um ATP no domínio GHKL para uma histidina do DHp. O destino desse fosfato é um aspartato de um domínio REC (abreviação do inglês para *receiver*). Se esse REC está presente em uma segunda proteína, essa é chamada cognata ou reguladora de resposta (RR). Cognata e HK compõem o chamado sistema de dois componentes (TCS - *two component system*). Quando a própria HK possui o domínio REC, a HK é chamada HK híbrida. Nesses casos, o grupo fosfato é transferido para um segundo REC, de uma outra proteína, através de uma proteína que contém o domínio HPt, abreviação derivada do inglês para *histidine phosphotransfer*. Em um segundo estado de uma HK ocorre a atividade fosfatase, caracterizado pela remoção do grupo fosfato do RR.¹²⁹

Já as MCPs são caracterizadas pelo domínio MA, abreviação para metil acceptor. Ele é composto de um feixe antiparalelo de quatro estruturas secundárias em hélices, chamadas MH (*methylation helices*).¹³⁰ Cada par de hélices vem de uma MCP, portanto, a função depende de dímeros de MCPs. Uma MCP depende de um complexo proteico de sinalização regulado por ela. No geral, esses complexos seguem a organização e a composição daqueles que regulam o flagelo em *E. coli*, que são usados como modelo. Nesses sistemas, cada uma das proteínas é nomeada com o prefixo Che, abreviação do inglês para quimiotaxia, *chemotaxis* (Figura 22A). Sítios específicos das MHs são metilados constantemente por CheR, o que faz o feixe de hélices assumir uma conformação fechada/compacta e apta para ativar a proteína CheA.

Com a ausência de nutrientes, ou a presença de repelentes, MCPs desencadeiam a ativação da CheA. Essa proteína é responsável por fosforilar os domínios REC de CheB e de CheY. Similaridades com vias de HKs são usadas para classificar o conjunto CheA e CheY como um sistema HK de classe II¹³¹ ou um sistema atípico de dois componentes.¹³² CheB fosforilada eleva sua atividade de remover grupos metil de sítios específicos nas MHs, tornando o feixe de MHs da MCP frouxo, incapacitando a ativação de CheA e, assim, desativando a sinalização. CheY fosforilada promove a mudança da rotação do flagelo de anti-horária (CCW, *counterclockwise*), regime no qual a célula é impulsionada em uma única direção, para horária (CW, *clockwise*). Com a rotação CW do flagelo, a célula se desloca randomicamente. Isso permite que a célula siga em frente apenas quando encontra um ambiente favorável. Por causa desse sistema, MCPs ativas e inativas podem ser referenciadas respectivamente como CW e CCW. Para voltar ao estado de propulsão em uma única direção, é necessário que CheZ

desfosforile CheY. Além disso, a atividade constante de CheR faz com que a MCP volte ao estado de inicial.¹³³

Figura 22 - Analogia entre sistema de sinalização clássico MCPs (A) e outro que inclui uma DGC (B).

A) A figura ilustra a organização celular dos componentes gerais de um sistema de sinalização coordenado por MCP. O estímulo geralmente são nutrientes, como serina, aspartato, glucose, oxigênio, etc. Experimentos com eletro crio-tomografia mostraram que MCPs formam redes na membrana, constituídos de clusters compostos por trimeros de dímeros.¹³⁴ Na figura, cada bastão e um círculo roxo representa um dímero de uma MCP. Redes de MCP podem ser formadas por sensores de diferentes moléculas.¹³⁵⁻¹³⁶ A disposição desses clusters está destacada no quadrado em preto. Para exemplificação, complexos LuxPQ, um sistema HK empregado no *quorum sensing*, não formam redes ordenadas como as observadas em MCPs¹³⁷; B) Distribuição dos componentes de proteínas que compõe o sistema de sinalização de proteínas codificadas no operon *wsp*. O análogo da CheY nesse sistema é a WspR, uma DGC, destacado com um quadro vermelho na figura. O c-di-GMP controla, dentre outros, o regulador transcricional FleQ. Ao ligar-se ao c-di-GMP, FleQ deixa de reprimir a produção de exopolissacarídeos Pel e expressão de genes flagelares. O estímulo da MCP desse sistema é a presença de uma superfície ou o contato com outra célula.

Fonte: Adaptada de SOURJIK; WINGREEN^{135,138}; PORTER; WADHAMS; ARMITAGE.¹¹⁶

Por volta de 10% de todas as DGCs atuam como reguladores de resposta. Por exemplo, a proteína WspA, uma MCP, desencadeia a fosforilação da WspR, uma DGC análoga da CheY do sistema de sinalização¹³² (Figura 22B). Outra MCP Tlp1 possui o domínio PilZ no fim da cadeia polipeptídica, mais próximo a região regulatória.¹³⁹ Exemplos de HK, como RetS, interfere na concentração de c-di-GMP intrabacteriana.¹⁴⁰ Um exemplo que chama a atenção a HK conhecida como QseC, que possui um domínio EAL entre os domínios DHp e CA.¹⁴¹

1.13 Regulação de domínios efetores

1.13.1 Mecanismos de captação de estímulos ambientais bacterianos

É possível montar um esquema geral representativo para organização de proteínas transdutoras de sinal bacteriano, partindo das semelhanças entre sistemas fosforilativos e da homeostase de segundos mensageiro (Figura 23).

Figura 23 - Modelo ilustrativo geral dos componentes básicos que podem aparecer em sistemas de transdução de sinal abordados no trabalho.

Sinais externos são captados a porção receptora (i), que em alguns casos requer o auxílio uma SBP ou mesmo possui o enovelamento de uma SBP. As mudanças estruturais geralmente são transmitidas por regiões ou domínios em hélices α (iv). Domínios regulatórios adicionais (iii) são encontrados auxiliando na transmissão ou combinando a sinalização com outras condições intracelulares para ativar ou desativar os domínios efetores (v).

Fonte: Elaborada pelo autor.

Os receptores de estímulos externos podem ser classificados de acordo com o local da captação do estímulo externo em relação a membrana. Isso foi proposto para HK¹⁴²⁻¹⁴³ e, de maneira similar, para MCPs.¹⁴⁴ Devido as várias relações desses sistemas com as rotas do c-di-GMP, é apresentado nesse trabalho uma versão simplificada dessas classificações (Figura 24A):

- Receptores citoplasmáticos, através de uma proteína ancorada ou não à membrana,
- Receptores embebidos na membrana;
- Receptores externalizados em relação ao citoplasma, meio extracelular ou periplasma, associados a proteínas transdutoras.

Figura 24 - Divisão e diversidade dos sistemas conhecidos de recepção de sinal ambiental.

A) Representação dos tipos de sinalização 1, 2 e 3. Os receptores do tipo 3 podem ser subdivididos ainda em α , compostos majoritariamente por α hélices, β , essencialmente compostos por fitas β e os $\alpha+\beta$, compostos por uma mistura de fitas e hélices; B) Os receptores do tipo mostrado em 3 em geral atuam como dímeros. Nas figuras uma unidade está em branco e a outra está em cores - hélices em ciano, loops em vermelho claro e fitas em magenta. Nos casos em que a captação é feita por complexos o dímero foi omitido por motivos de clareza. As PBP (ou SBPs) estão coloridas em laranja. Os N-terminais (N) estão destacados em azul escuro, os C-terminais (C) em vermelho e as regiões do ligante estão marcadas com uma estrela. Os PDBs das estruturas mostradas são TorS-TorT: 3O1H; Tar-aspartato: 1VLT; NarX-nitrato: 3EZH; CitA-citrato: 1P0Z; PhoQ-Ni²⁺: 3BQ8; LuxQP-AI-2: 1ZHH; RetS: 3JYB.

Fonte: Elaborada pelo autor.

A sinalização por módulos externos ao citoplasma é mais numerosa em bactérias e possui diversos exemplos já caracterizados.¹⁴⁵ Nesse grupo, a região proteica na qual ocorre a captura do sinal é geralmente delimitada por hélices transmembranares. Esses grupo de receptores podem ser novamente divididos em três classes, baseadas em classificações propostas em trabalhos do grupo de Hendrickson e colaboradores.^{143,146} Essa divisão está de acordo com a composição de estruturas secundárias em enovelamentos já observados (Figura 24B):

- compostos por mistura de hélices e fitas, $\alpha+\beta$ e α/β ;
- puramente constituídas por α -hélices (abreviado como α);
- predominantemente constituídos por fitas β (representado por β);

Em todos exemplos de receptores externalizados é observada a atuação em dímeros. Um ou dois sítios do ligante por dímero já foram observados para diferentes exemplos. Em alguns casos, há a necessidade ainda um cofator, como grupo heme,¹⁴⁷ ou um parceiro para recepção

do sinal. Esses parceiros de interação são proteínas ligadoras de substrato (*substrate-binding proteins* – SBP),¹⁴⁸ encontradas também como proteínas ligadoras periplasmáticas (*periplasmic binding protein* – PBP). Vários exemplos mostraram que a interação com o ligante ou o parceiro de interação promove, na totalidade da porção voltada para fora do citoplasma, a alternância entre conformações assimétrica e simétrica^{129,149-150}

As SBPs possuem um domínio característico SBD, que é extremamente versátil pois aparece também em receptores externos¹⁵¹ com enovelamento já referenciado como α/β , ou PBP-like.¹⁴³ Além disso, SBPs são mais reconhecidas por atuarem juntamente com importadores de membrana ABC (*ATP-binding cassette*).¹⁴⁸ Em comum, todas as SBPs são compostas por duas porções que geralmente se aproximam na presença do ligante. Essas proteínas já foram categorizadas em até três classes (I, II e III) de acordo com similaridades estruturais e composição de fitas beta e hélices alfa.

Já o grupo de receptores $\alpha+\beta$ são representados pelo enovelamento PDC, que possui similaridade estrutural com o domínio PAS, clássico de receptores (seção 1.13.2, a seguir). Por isso, esse domínio já foi referenciado como PAS, PAS-like ou PDC/PAS.¹⁴⁵ A abreviação PDC vem das três proteínas sensoras nas quais foi primeiramente caracterizado: PhoQ, DcuS e CitA. PhoQ também é uma HK ativa em baixas concentrações de cátions divalentes, baixo pH e por peptídeos antimicrobianos catiônicos.¹⁵²⁻¹⁵³ CitA, DcuS além de DctB,¹⁵⁴ outro receptor PDC vastamente citado na literatura, são HKs, com ligantes dicarboxilatos de 3 a 6 carbonos, como citrato, aspartato, fumarato, malonato e succinato.¹⁵⁵ Outro exemplo de receptor $\alpha+\beta$ é LuxQ, HK que sinaliza a presença do autoindutor 2 (AI-2^j), sinalizador do *quorum sensing*.

Nas estruturas dos receptores de LuxQ e DctB, assim como em outros exemplos,¹⁴⁶ são encontrados dois enovelamentos PDC em série, um membrano distal e outro proximal. A interação desses chamados *double-PDC* com o ligante é observada comumente no PDC membrano distal.¹⁴³ Enquanto a interação com o ligante é direta em DctB, LuxQ precisa como intermediador uma SBP chamada LuxP. LuxP, ao interagir com a molécula AI-2, inativa a atividade HK de LuxQ.¹⁵⁶

Para receptores α , um exemplo clássico é Tar (MCP), que se mantém inativa ao se ligar diretamente a aspartato ou maltose, através da proteína ligadora de maltose MBP (*maltose binding protein*). Tar ainda é ativa na presença de Co^{2+} e Ni^{2+} .¹⁵⁷ NarX também é um receptor α , que é ativa na presença de nitrato ou nitrito e possui um único sítio por dímero, diferente da

^j AI-2 faz referência a molécula (2S, 4S)-2-metil-2,3,3,4-tetrahidroxitetrahydrofurano-borato, ou simplesmente S-THMF-borato.²⁵

Tar, com um sítio membrano distal em cada monômero. Outro exemplo TorS, HK que sinaliza a presença do N-óxido de trimetilamina (TMAO), opção de receptor de elétrons da cadeia respiratória na ausência de oxigênio. A ligação ocorre com auxílio da PBP denominada TorT. A estrutura de TorS pode ser dividida em duas unidades em hélices, uma membrano distal e outra membrano proximal, por isso essa estrutura foi referenciada como *double all- α* .¹⁴³

RetS é exemplo único já descrito de sensor periplasmático compostas predominantemente por fitas beta. É possível encontrar a porção sensora também com o enovelamento de SBPs, também composto por fitas e folhas.¹⁵⁸ Não se conhece o ligante da RetS (HK com PDB do receptor 3JYB), mas sabe-se que RetS regula negativamente a produção de c-di-GMP.¹⁴⁰

1.13.2 Estrutura e função dos domínios com enovelamento PAS

O domínio PAS é encontrado em todos os reinos da vida compondo um extremo número de proteínas registradas em bancos de dados de sequências. PAS é a sigla para três proteínas de *Drosophila melanogaster*, nas quais foi primeiramente identificado, com siglas PER, ARNT e SIM. Esse domínio pode ser encontrado com vários outros nomes anteriores como repetição PAS, S1/S2 box, PAS/PAC ou PYP-like.¹⁵⁸ O domínio PAS, assim com seu enovelamento, é empregado em uma diversidade de receptores, tanto em receptores do citoplasma ou, com visto, em receptores $\alpha+\beta$. Esses receptores captam uma diversidade de estímulos, como biomoléculas, gases, potencial redox e luz. O domínio é encontrado ainda em diferentes vias de sinalização, inclusive nas vias do c-di-GMP. A importância do domínio para as vias do c-di-GMP vem da recorrência do mesmo nas vias do c-di-GMP: a sequência de domínios PAS-GGDEF e PAS-GGDEF-EAL são as mais numerosas anotadas em bancos *online*.¹¹⁷

O domínio PAS possui aproximadamente 100-120 resíduos, composto por uma folha antiparalela de topologia 2-1-5-4-3¹⁵⁹ com geralmente mais 4 hélices de tamanho variável distribuídos em um dos lados da folha, formando um bolsão. Essas fitas e hélices são enumeradas sobre uma numeração única, de A a I (Figura 25).

Apesar da variedade de ligantes, a maioria das interações do domínio PAS com seus ligantes ocorre na face interior da folha e com as hélices E α e F α . Além disso, o loop FG possui resíduos chave para a interação, sendo observado sob uma conformação ordenada quando ligado.¹⁵⁶ São poucos os aminoácidos conservados, apesar da conservação do enovelamento. Eles incluem resíduos hidrofóbicos nos limítrofes das estruturas secundárias e uma tríade de resíduos DIT no final do domínio. O aspartato (D) da tríade geralmente interage com uma lisina

ou arginina do loop entre as fitas G e H. O último resíduo da tríade (T), por sua vez, realiza uma ligação hidrogênio com um grupo amina da cadeia principal. Foi proposto que o motivo DIT é responsável por transmitir a mudança conformacional do núcleo do domínio para as estruturas secundárias adjacentes.^{128,159}

Figura 25 - Enovelamento PAS.

A) As estruturas secundárias do PAS são etiquetadas como definido para FixL. Ela tem a função de ativar seu cognato FixJ na hipóxia, desencadeando a fixação de nitrogênio¹⁶⁰⁻¹⁶¹; B) Variações na sequência de estruturas secundárias podem ocorrer.; C) O enovelamento, ilustrado pela FixL (PDB 1dp6), por vezes, já foi comparada na literatura como uma luva de beisebol.¹⁶⁰⁻¹⁶²

Fonte: Elaborada pelo autor.

Em muitos exemplos descritos, o domínio PAS tem papel na dimerização, ou através de resíduos apolares na face exposta da folha ou, como mais comum em proteínas de procariotos, através das hélices flanqueadoras do domínio, a N-cap¹⁵⁹ e Jα.¹⁶³ Em alguns casos a função do domínio PAS é apenas para oligomerização, como em PpsR¹⁶⁴ de *Rhodobacter sphaeroides* (PDBs 4l9e, 4l9f e 4l9g), em RbdA¹¹⁹ de *P. aeruginosa* (PDBs 5xgb, 5xgd e 5xge, Figura 26A) ou na proteína KinA,¹⁶⁵ de *B. subtilis*. RbdA possui tanto atividade DGC quanto PDE e abrange os domínios PAS-GGDEF-EAL.¹¹⁹ A função do domínio PAS de VicK de *Streptococcus mutans* não é conhecida, mas sabe-se que forma um dímero estabilizado por um zíper de leucina, formada na hélice N-terminal do PAS (PDB: 4i5s)¹⁶⁶ (Figura 26B).

O domínio PAS é comumente copurificado com ligantes de sistemas de expressão heterólogos. Um exemplo vem com DosP, uma fosfodiesterase detentora de dois domínios domínio PAS. Desses dois domínios, o N-terminal é ligador de grupo heme. DosP possui também um domínio GGDEF inativo e um EAL ativo. Ela é copurificada em um complexo ribonucleoprotéico que inclui a DGC DosC. Picos de absorbância característicos de cofatores podem ser usados como primeiro indício da sua presença. Um dos exemplos é AxDGC2, que

contem o domínio PAS que liga um grupo flavina oxidado (FMN ou FAD). Na purificação da proteína é relatado uma cor amarelada da amostra devido ao cofator. Por exemplo, o grupo flavina possui máximos de absorção nos comprimentos 375 nm, 450 nm e 470 nm.¹⁶⁷ Domínios PYP, que possuem enovelamento PAS, apresentam copurificação com ácido p-hidroxicinâmico (HCA), ou p-cumárico, ligante com absorbância característica no 400 nm.¹⁶⁸⁻¹⁶⁹ Em outro exemplo, no domínio PAS de Aer2 de *P. aeruginosa* foi possível monitorar a presença do grupo heme observando seus máximos de absorbância no 418 nm, 542 nm e 577 nm.¹⁷⁰

Figura 26 - Estruturas RbdA (A) e VicK (B), exemplo nos quais o PAS é empregado para dimerização. RbdA é um exemplo de estrutura na qual a dimerização ocorre pelo hélice Jα com uma das hélices de dimerização do EAL; B) Estrutura de VicK, na qual a dimerização ocorre pela hélice N terminal do domínio PAS, denominada N-cap.

Fonte: Elaboradas pelo autor.

O enovelamento PAS é encontrado em outros domínios, referenciados como *PAS-like*. Alguns exemplos são dados com o domínio LOV, que responde a luz, oxigênio e voltagem por meio dos cofatores FMN ou FAD,¹⁶³ Cache, encontrado em canais de cálcio e receptores quimiotáticos (do inglês, *chemotaxis*),¹⁵⁹ YkuI_C, domínio C-terminal da proteína YkuI,¹⁰⁶ ou em domínios PYP, de proteínas fotoreativas ao amarelo, do inglês *photoactive yellow protein*, ou correlatas (*PYP-like*).¹⁶⁹ Em um trabalho de Zhulin e colaboradores foi proposto uma unificação dos domínios sensores externos com enovelamento PAS ao outro grupo denominado Cache.¹⁴⁵

O PAS ainda possui relações com o domínio GAF, acrônimo das proteínas nas quais primeiro foi encontrado: em fosfodiesterases específicas de GMP cíclico, adenilato ciclases e no ativador transcricional FhlA.^{145,171} O mesmo tipo de folha encontrada no PAS é observada no domínio GAF, com o acréscimo de uma fita após B β (fita 2), alterando a topologia para 3-2-1-6-5-4. Além disso, o GAF pode possuir algumas hélices extras entre as fitas.¹⁵⁹ PAS e GAF compõe o grupo de proteínas com enovelamento porfirina.¹⁷² Semelhante ao PAS, o GAF é reconhecido amplamente por se ligarem aos sinalizadores cíclicos cGMP e cAMP, mas também por ser receptor de uma série de outros estímulos.¹⁷³

1.13.3 Domínios CHASE

Os domínios CHASE aparecem vastamente como receptores externalizados em relação ao citoplasma. A importância desses domínios vem do fato que são sensores que aparecem em elevado número em diversos sistemas de sinalização, não apenas naqueles compostos por DGCs e PDEs, mas também em outras ciclases e quinases. Os domínios CHASE desempenhariam a função de receptores extracelulares, por isso a abreviação do inglês para *cyclase/histidine kinase-associated sensing extracellular*.⁶ Consistem em 9 grupos independentes enumerados de 1 a 9 (CHASE, CHASE2-9). Das aproximadamente 400 000 sequências nas quais são identificados os domínios GGDEF, EAL ou HD-GYP registradas no Uniprot¹⁷⁴, 2,5% possuem algum dos domínios CHASE (Figura 27).

Cada CHASE foi originado através de buscas de sequências pelo PSI-BLAST utilizando o banco de sequências não redundantes (nr) do NCBI. Essas buscas eram iniciadas com a porção predita receptora de sinal externo, flanqueada por hélices transmembrana e de proteínas sinalizadoras distintas.⁶ Esses trabalhos iniciais foram feitos por Zhulin⁴ e Aravind⁵ e colaboradores. Os CHASE7-9 foram definidos em um trabalho recente de Hengge e colaboradores, baseadas nas regiões periplasmáticas de DgcQ (YedQ), DgcN (YfiN), e PdeI (YliE).⁷ Nesse mesmo trabalho foram introduzidas as classificações dos domínios GAPESs, sigla do inglês para sensores periplasmáticos de gamaproteobactérias (*gammaproteobacterial periplasmic sensor domain*), domínio encontrado na proteína STM3615.

Figura 27 - Distribuição de domínios das vias do c-di-GMP nos grupos CHASE.

Gráfico com alguns números de seqüências CHASE anotadas no banco Uniprot, com destaque para aquelas também com os domínios GGDEF, EAL, GGDEF-EAL e HD-GYP.^k

Fonte: Elaborada pelo autor.

Dentre os CHASE, é predito que CHASE e CHASE4 assumam o enovelamento PDC^{145,175} (Figura 28C). O CHASE4 foi definido através de buscas por similaridade iniciadas com o receptor periplasmático de uma HK de *Archaeoglobus fulgidis* (gi 11499310; número de acesso NCBI AAB89530). O domínio CHASE4 possui a assinatura conservada WWD, de função desconhecida (Figura 28A).

Já para CHASE3, existe uma única estrutura resolvida do domínio (PDB 3VA9) com trabalho vinculado ainda a ser publicado. O CHASE3 possui uma estrutura em α hélice (Figura 28B), que se assemelha ao receptor do sensor de aspartato Tar (Figura 28D). A sequência original usada para definir o CHASE3 foi a porção periplasmática CyaA, adenilato ciclase (AC)¹⁰⁵ que monitora estresse osmótico. Esse domínio é caracterizado pelo motivo RGFLLT, o qual também não se conhece a função.

^k Códigos Pfam dos domínios: GGDEF:pf00990; EAL: pf00563; HD-GYP: pf01966; CHASE: pf03924; CHASE2: pf05226; CHASE3: pf05227; CHASE4: pf05228; CHASE5: pf17149; CHASE6_C:pf17150; CHASE7: pf17151; CHASE8: pf17152; CHASE9: pf17153.

Figura 28 - Exemplificação das arquiteturas secundárias e terciárias de domínios CHASE, ilustrados pelos CHASE e CHASE4 (enovelamento PDC) e pelo CHASE3.

A) Composição de hélices e fitas da proposta original para o CHASE4. A sequência de resíduos mostrada WDD é o padrão de conservação que mais se destaca no domínio; B) Composição de estruturas secundárias preditas segundo trabalho original que definiu o domínio CHASE3, com destaque para o motivo característico RGLLLT; C) Estrutura da proteína DcuS, de *E. coli*, ilustrando o enovelamento PDC, mesmo predito para CHASE e CHASE4; D) Alinhamento estrutural é entre a fração sensora de Tar (domínio TarH)¹⁷⁶ e o CHASE3 resolvido (RMS:3,4 Å);

Fonte: B) e C) Adaptada de ZHULIN; NIKOLSKAYA; GALPERIN⁶; A), D), E) Elaborada pelo autor.

1.13.4 Domínio HAMP

É extremamente comum a presença do domínio HAMP em proteínas transmembrana de sistemas de sinalização de procariotos e alguns eucariotos. O HAMP se encontra em proteínas transdutoras de sinal, sendo a abreviação para HK, AC, MCP e algumas fosfatases (*phosphatases*).¹⁰⁵ O domínio geralmente disposto próximo a face citossólica da membrana em proteínas transdutoras. Em modelos de sinalização, o domínio funcionaria como uma chave de liga e desliga, que processa e responde a informação proveniente das hélices transmembrana para os domínios de resposta. Mas a maneira como ele atua ainda está em debate.

O domínio HAMP é composto por duas hélices anfipáticas (AS1 e AS2), interligadas por uma região conectora (CTR) desestruturada. O padrão característico de *coiled-coil* é presente nas hélices do domínio (Figura 29A).

Figura 29 - Elementos do domínio HAMP, sequencial e tridimensional (A) e (B), e os mecanismos descritos para os domínios em (C), (D) e (E).

A) O domínio HAMP é composto por duas hélices anfipáticas, denominadas AS1 e AS2, e uma região conectora que estabiliza o domínio. Regiões adjacentes, como o cabo regulador e a alternadora de fase estão associadas com a captação e transmissão da mudança conformacional na sinalização celular; B) O domínio funcional conhecido é formado por um dímeros enovelados de duas formas. O HAMP(A) é um *coiled-coil* tetramérico *x-da*, e o HAMP(B) é um feixe distinto, com as hélices AS1 afastadas e AS2 próximas na região C-terminal. Outra diferença de HAMP(B) para HAMP(A) é que as hélices AS2 se encontram deslocadas para região N-terminal; C) A primeira estrutura do HAMP Af1503 possui enovelamento HAMP(A). A possibilidade de converter o *coiled-coil x-da* para o clássico *a-d* fundamentou o primeiro mecanismo de transmissão da sinalização, que envolveria rotações das hélices, como mostrado na figura; D) o enovelamento HAMP(B) inspirou o modelo de hélices dinâmicas e o de hélices tesoura.

Fonte: A) e B) Elaboradas pelo autor; C) Adaptada de SCHULTZ; KANCHAN; ZIEGLER¹⁷⁹; D) Adaptada de AIROLA et al ¹⁸⁹; E) Adaptada de STEWART.¹⁷⁷

Foram observadas na forma de dímeros, formando um feixe de hélices tetramérico paralelo. Dentre os resíduos comuns em HAMPs está uma glicina, altamente conservada e necessária para acomodar uma volta acentuada no final da hélice AS1 e CTR. O motivo conservado DEXG no início de AS2 formam um anel de interação com uma prolina conservada de AS1. Trabalhos sugerem que o motivo DEXG está diretamente envolvido na transmissão de sinal recebido pelas hélices transmembrana.^{131,183} O motivo DEXG geralmente é ausente em HAMPs que estão em múltiplas cópias consecutivas em uma mesma sequência.

Em Aer, MCP de *E. coli*, é proposto que o domínio PAS transmite a sinalização através de contatos com a hélice AS2 do HAMP. Aer é sensor de oxigênio, indiretamente, através da ligação a FAD. Essa ligação é feita por um domínio PAS internalizado, ao contrário de outras MCPs clássicas de *E. coli* que a porção receptora do sinal é externalizada.¹⁷⁸⁻¹⁷⁹

A defasagem entre *coiled-coils* em regiões que conectam domínios promovem a alternância de estados conformacionais entre esses domínios. Essas regiões de $7n+3$ ou $7n+4$ são denominadas em trabalhos com HAMP por alternadores de fase (tradução do inglês para *phase stutter*).¹³⁰ Estudos sugerem ainda que a região conectora entre a AS1 e a hélice transmembrana adjacente é a responsável por ajustar a tensão no empacotamento do HAMP, manipulando sua estabilidade (Figura 29A).¹⁸⁰⁻¹⁸¹

Três linhas explicam a maneira como o HAMP funciona, baseadas em evidências experimentais. A primeira teve como fundamento a estrutura pioneira de Af1503 de *Archaeoglobus fulgidus*, proteína curiosamente sem domínio efetor na mesma cadeia polipeptídica. O feixe de quatro hélices do HAMP forma um *coiled-coil* que não segue a montagem clássica *kobs-in-holes* ou *a-d* mas outra, denominada *knobs-in-knobs* ou *x-ad*. Rotacionando o empacotamento *x-da* em aproximadamente $+26^\circ$ na AS1 e -26° na AS2 se obtém a conformação *a-d* (figura 29C).¹ A possibilidade dessa conversão culminou no primeiro mecanismo de atuação do HAMP, ou modelo de engrenagens (*gearbox*). Nele, cada conformação de *coiled-coil* corresponde a um estado da proteína sensora.

Com a resolução atômica dos três primeiros HAMPs N-terminais e em tandem da proteína Aer2 de *P. aeruginosa* (PDB: 3lnr) foi fundamentado outros dois mecanismos de transmissão. Aer2 é um sensor citoplasmático que possui cinco domínios HAMP com um domínio PAS ligador de grupo heme inserido entre o terceiro e o quarto HAMP.¹⁸² Desses 3 domínios com estrutura resolvida, o primeiro e o terceiro HAMP se assemelha ao enovelamento

¹ A repetição característica de 7 resíduos em *coiled-coils* corresponde a duas voltas para cada hélice. Cada resíduo da repetição ocupa uma posição característica, distantes entre si por $\approx 51,4^\circ$ ($360^\circ \div 7$). Portanto, uma rotação de $\approx 25,7^\circ$ nas hélices desloca o padrão característico para posições intermediárias.

encontrado em Af1503, que será referenciado aqui como HAMP(A). Enquanto isso, o segundo HAMP (HAMP2) não possui o pacote *x-da* nem o *a-d*, mas uma terceira conformação, referenciada como HAMP(B).¹³⁰ No HAMP(B) as porções N-terminais do feixe se encontram afastados e as porções C-terminais se encontram próximas. Na porção C-terminal as hélices assumem uma conformação de um *coiled-coil* dimérico.¹⁸³ Além disso, a conversão de HAMP(A) para HAMP(B) requer um movimento da hélice AS1 no sentido C-terminal do domínio.¹⁸⁴ Isso é contundente com o movimento pistão de TM2 para o interior da célula. Duas posições conservadas estão na região conectora e são ocupadas por resíduos hidrofóbicos importantes para o empacotamento de HAMP(B), HR1 e HR2 (*hydrophobic region*).^{130,184}

No modelo de hélices dinâmicas (*dynamic bundle*), propõe que as conformações HAMP(A) e de HAMP(B) são intercambiáveis e imprimem a mesma resposta nos domínios efetores. Enquanto isso, estados intermediários entre essas conformações imprimem a forma inativa no efector. Já no modelo de hélices tesoura (*scissoring helices*) cada uma das conformações corresponde a um estado do sensor.^{185,192} Nesses dois últimos modelos é possível encontrar por vezes o HAMP(A) chamado de feixe justo (*tight bundle*), conformacionalmente heterogêneo, segundo experimentos de ressonância paramagnética eletrônica (EPR). Enquanto isso, o enovelamento de HAMP(B) possui o feixe frouxo (*loose bundle*), conformacionalmente homogêneo segundo experimento de EPR¹³⁰ (figura 29). Trabalhos com MCPs mostram que é possível obter proteínas em diferentes estados, ativos ou inativos, fazendo modificações no domínio HAMP. Essas alterações englobam tanto mutações pontuais^{130,183,186-187} ou remoção de partes do domínio.¹⁸⁸

Apesar da forma dimérica recorrente em exemplos de proteínas com HAMP, a porção contendo HAMP-GGDEF de YfiN (YfiN^{HAMP-GGDEF}) se revela quase que exclusivamente no estado monomérico. No trabalho de Giardina e colaboradores, YfiN se revelou em solução 98% monomérica, segundo experimentos de sedimentação por ultracentrifugação. Então, para explicar a atividade de YfiN^{HAMP-GGDEF}, foi proposto um estado de transição rápido e dimérico.¹¹¹

1.13.5 Transmissão da sinalização através de mudanças conformacionais por hélices

A transmissão de sinal entre domínios através da mudança conformacional intermediada por hélices α é um tema extremamente recorrente em sistemas de sinalização bacterianos. Domínios regulatórios como HAMP, hélice-S¹³⁷ e hélice-C¹⁸⁹ são essencialmente compostas por hélices α e comumente conectam domínios. Outro exemplo vem de extensões em hélices

na porção N-terminal e C-terminal de domínios PAS. Para domínios LOV, são relatados diferentes mecanismos de como a hélice C-terminal denominada Jα transmite a sinalização, desencadeada por luz (Figura 30A).

Figura 30 - Exemplos da transmissão de sinal pela hélice C-terminal Jα do domínio LOV, que possui enovelamento PAS (A) e padrão hidrofóbico coiled-coil comumente conectando PAS e GGDEFs (B).

Fontes: Adaptada de MINASOV et al ¹²⁹;CHRISTEN.¹³⁰

Hélices transmissoras de sinal podem ainda compor chamado *coiled-coil* (CC), enovelamento exclusivamente composto hélices. A identificação de *coiled-coils* são facilitadas por causa de um padrão característico de repetição de aminoácidos hidrofóbicos. E esse padrão é vastamente encontrado compondo o N-terminal de DHps¹⁹⁰, de MCPs e GGDEFs^{97,191} (Figura 30B).

Outros exemplos de extrema relevância são os três mecanismos propostos de transdução de sinal para o citoplasma através das hélices transmembrana. Num deles ocorre o deslocamento relativo entre as hélices perpendicular a membrana, chamado deslocamento pistão (*piston*). Numa outra forma de transmissão as hélices se afastam paralelamente a membrana, movimento denominado tesoura (*scissor*). Numa terceira forma de transmissão as hélices rotacionam em relação ao próprio eixo (*rotating helix*) Figura 31.

Figura 31 - Tipos de mudanças conformacionais transmitidas por porções sensoras propostas.

No geral, a mudança do estado assimétrico e assimétrico, ou vice-versa, do receptor, promove deslocamentos relativos das hélices que ancoram a região sensora. No modelo pistão (*piston*) o deslocamento pode ser para o interior da célula, como nos receptor α de MCP Tar ou Tsr na presença de moléculas atrativas (nutrientes), ou para o exterior, como nos receptor α de HK NarX na presença de nitrato ou nitrito ou Tar e Tsr na presença de repelentes.¹³⁰ Esse modelo é suportado diferentes técnicas. Uma delas é através pontes dissulfeto artificiais, que fixam conformações. A segunda é de dados estruturais de receptores apo e ligados. A última vem de resultados com a substituição de resíduos embebidos na membrana.¹⁹² O segundo mecanismo, de movimento tesoura, culminou de estudos de *crosslinking* com a porção sensora da PhoQ.¹⁵²⁻¹⁵³ O terceiro modelo, envolve a rotação hélices, é baseado nas formas apo e ligada do complexo LuxQ-LuxP com e sem o ligante, o autoindutor 2 (ou AI-2).

Fonte: Adaptada de BHATE et al.¹⁶²

1.14 Sistemas estudados e objetivos do trabalho

Baseado no estado da arte do tema e motivados pela importância do mesmo, foram formuladas as seguintes perguntas gerais do presente trabalho, que são:

- Como sinais externos ativam enzimas transmembrana que regulam os níveis de c-di-GMP intracelulares?
- Como os sinais externos afetam a atividade catalítica e/ou as interações em complexos enzimáticos?

- Como proteínas, tanto com domínio GGDEF quanto EAL, ativos ou não, são reguladas ou exercem função regulatória?

Para abordar essas questões, no presente trabalho foram estudadas proteínas contendo GGDEF e EAL e que fossem transmembranares. Destaque foi dado para domínios proteicos periplasmáticos CHASE, pela elevada presença nas vias do c-di-GMP e em outras. Outro tema inserido no trabalho foi trazer algumas pistas de como é a dinâmica entre domínios GGDEF e EAL em tandem. Para isso, aplicou-se buscas e análises de bioinformática utilizando, entre outras, a transdutora de sinal STM3615.

1.14.1 Revisão sobre as proteínas do trabalho

As proteínas são da cepa *P. aeruginosa* UCBPP-14 (PA14), uma delas codificada no gene PA14_53310. Essa proteína nesse texto é referenciada pelo código do gene do ortólogo de *P. aeruginosa* PAO1, PA0847. Isso devido à alta identidade entre as proteínas ortólogas além do fato que a cepa PAO1 é usada comumente como referência na literatura.^m

Além disso, foram estudadas outras proteínas de *P. aeruginosa*, pois interagiram com PA0847 em experimentos de duplo híbrido¹³, podem então desempenhar ou compor algum papel de sinalização conjunta com PA0847. Esses parceiros de interação são as frações N-terminais das proteínas codificadas nos genes PA14_20820 e PA14_40570 e a proteína inteira PA14_55160. Essas proteínas possuem ortólogos de PAO1 codificados nos genes PA3343, PA1851 e PA0707.ⁿ PA3343 é descrita como HsbD⁸, apresentou interação forte com PA0847, segundo resultados prévios, e por isso foi selecionada para o estudo. PA1851 é bem similar a HsbD e então foi usada para confrontar resultados. Como a interação é um dos mecanismos utilizados para alcançar a especificidade em rotas de sinalização por c-di-GMP, foi proposto que PA0847 e seus parceiros de interação citados possam compor um desses sistemas específicos.^{13,16}

PA0847 possui uma fração voltada para o periplasma, a qual abrange um domínio CHASE4, e uma fração citoplasmática, na qual são identificados os domínios HAMP, PAS e GGDEF. PA1851 possui a mesma arquitetura de domínios da HsbD. Ambas possuem uma

^m Uma das diferenças entre essas cepas é que PA14 não possui a DGC PA2771 e a PDE PA2818, também chamada de arr. Em contrapartida, a cepa PA14 possui a PDE pvrR, ausente em PAO1.¹⁹³

ⁿ PA0707 é encontrada referenciada como RegA¹⁹⁴ ou ToxR¹⁹⁵, proteína que interfere na expressão da exotoxina A (EXA), considerada o fator de virulência extracelular mais poderoso de *P. aeruginosa*.¹⁹⁴ Como essa proteína apresentou uma interação fraca com PA0847 e com diversas outras proteínas, não foi dado para a mesma.

fração N-terminal embebida no citosol de 51 resíduos e uma região seguida por 6 hélices transmembrana (TM), seguida de um domínio GGDEF conservado. Dentre os parceiros de interação, um foco maior foi dado para porção solúvel e N-terminal de HsbD, por trazer um resultado de interação forte com a fração citossólica de PA0847, ao contrário dos outros parceiros que forneceram uma interação fraca, como indicaram dados de duplo híbrido.¹³

Figura 32 - Arquitetura ilustrativa de PA0847, das proteínas que possuem interação com a fração citossólica da mesma (PA0847-HPG-1) e de STM3615.

PA0847 interage com PA0707 e a porção N-terminal de HsbD (HsbD-Nt) e de PA1851 (PA1851-Nt). *TM: hélice transmembrana.

Fonte: Elaborada pelo autor.

PA0847 e HsbD (PA3343) promovem altos níveis de c-di-GMP quando superexpressas.¹⁹³ Quando HsbD e PA0847 são deletadas/mutadas do genoma^{193,196}, a formação do biofilme não se altera e PA0847 eleva consideravelmente os níveis de motilidade da bactéria.¹⁹⁷ HsbD eleva a quantidade total de c-di-GMP quando superexpressa, sem promover com isso a formação de biofilme. Ademais, frente a certas condições testadas, a deleção do genoma da HsbD leva ao aumento da formação do biofilme. Em outro trabalho com *P. fluorescens*, também foi mostrado o impacto negativo de proteínas GGDEF sobre a formação do biofilme, o que suporta que o observado para HsbD não se trata apenas de um artefato experimental.⁹ Em contrapartida, quando PA1851 é superexpressa, não há alteração do c-di-GMP intracelular, nem formação de biofilme (Tabela 3 e Figura 33).

Tabela 3 - Dados da literatura de superexpressão de determinadas DGCs.

Cepa	c-di-GMP intracelular
PA14 (controle)	x1
++PA0847	x16
++PA1851	x1
++PA3343	x22

Fonte: Adaptada de KULESEKARA et al.¹⁹³

Figura 33 - Reprodução de alguns dos resultados de análise de biofilme das cepas com superexpressão (marcados com ++) e deleção (marcados com -) de genes envolvidos no presente trabalho.¹⁹³

Fonte: Adaptada de KULESEKARA et al.¹⁹³

HsbD é a abreviação para *Hpt-B-dependent secretion and biofilm*, por interferir no biofilme de maneira dependente da cascata de sinalização controlada por HptB. Dados apontam que HsbD é localizada no polo bacteriano, desaparecendo apenas no polo celular da célula filha durante a divisão celular (Figura 34A). HsbD interage com a forma fosforilada de HsbA. HsbA fosforilada, eleva a produção de c-di-GMP e do sRNA (*small RNA*) chamado RsmY. Esse sRNA sequestra a proteína ligadora de mRNA, RsmA, e a ligação reprime a transcrição da SadC, uma das DGC com maiores atividades de *P. aeruginosa* (Figura 34B). Para explicar a fuga da relação entre c-di-GMP e a formação do biofilme foi levantada a hipótese de que, uma vez que HsbD é localizada no polo da célula, apenas os receptores daquela região serão afetados.⁸

Figura 34 - Esquemática de dados da literatura sobre a proteína HsbD.

A) HsbD se encontra no polo celular de *P. aeruginosa*. Durante a divisão celular ela desaparece do polo celular da célula filha, surgindo novamente depois de concluída a divisão celular. B) HsbD é uma das proteínas da via da proteína HptB. A interação entre HsbA no estado fosforilado elimina a ação inibitória sobre a ação de SadC, outra DGC, o que induz a formação de biofilme.

Fonte: Adaptada de VALENTINI et al.⁸

Outra abordagem do trabalho é o estudo da dinâmica entre domínios GGDEF e EAL em tandem. Para isso, aplicou-se buscas e análises de bioinformática utilizando, entre outras, a transdutora de sinal STM3615. A importância dessa proteína vem de estudos de deleção gênica, que mostraram que a ausência dessa proteína compromete severamente a sobrevivência de *S. enterica* em camundongos infectados e o impacto da deleção destoa de qualquer outra deleção analisada.¹⁹⁸ É predita que essa proteína possui um domínio GAPES3, seguido dos domínios citoplasmáticos HAMP, GGDEF e EAL, nesta ordem na cadeia polipeptídica (Figura 35). No trabalho de doutorado de Flávio Rosseto, foi constatada a atividade EAL *in vitro* e a estrutura da porção GGDEF-EAL, que abrange os resíduos 242 a 668.¹⁶ Ao contrário de outros EALs já descritos, o domínio EAL de STM3615 a dimerização é independente do ligante e de outro domínio. Além disso, ele é um exemplar único de estrutura com GGDEF não funcional cataliticamente, seguindo de um EAL catalítico. Outro resultado anterior veio de comparações com outras proteínas GGDEF-EAL, que permitiu constatar que a hélice que conecta os domínios GGDEF e EAL por ser maior em proteínas DGC e/ou PDE é a responsável conferir uma rigidez entre domínios que, à princípio, permitiriam uma alternância de estados entre os domínios.

Figura 35 - Arquitetura ilustrativa da distribuição de domínios de STM3615.

*TM: hélice transmembrana.

Fonte: Elaborada pelo autor.

1.14.2 Estratégias experimentais

Para entender o funcionamento de PA0847 e o papel da interação com a região N-terminal de HsbD, e PA0851, foi proposto aplicar técnicas em biologia molecular para produzir as proteínas abordadas. A obtenção das proteínas então viabilizaria caracterizações biofísicas, com protocolos baseados em dados da literatura e outros, já usuais no laboratório do grupo de Biofísica Molecular “Sérgio Mascarenhas” (IFSC-USP), parte deles adaptados do material de referência.¹⁹⁹ Esses experimentos incluem:

- Entender sobre as formas de oligomerização das amostras por cromatografia de exclusão molecular (SEC) ou por cromatografia de exclusão acoplada a espalhamento de luz (SEC-MALS), dada a importância já conhecida das formas oligoméricas para regulação de proteínas das vias do c-di-GMP;
- Analisar a atividade de síntese ou degradação do c-di-GMP, por HPLC e/ou por *kit* comercial EnzChek;
- Cristalizar e obter mapas de coordenadas atômicas das construções das proteínas, a fim de obter informações estruturais que fundamentariam hipóteses sobre a função.

Para a investigação da interação inédita entre fração citossólica de PA0847 e a região N-terminal de HsbD (HsbD-Nt), foi proposto:

- A investigação da relevância biológica de HsbD-Nt, através de análises de conservação de sequências;
- Aplicar a técnica de *pull-down*, teste que ofereceria suporte para interação;

- A reprodução de experimentos de duplo híbrido bacteriano, acrescentando experimentos com diferentes partes de PA0847 contra HsbD-Nt, com a finalidade de mapear a interação em PA0847, ou seja, localizar em qual domínio ocorre a interação;
- Ensaios por titulação calorimétrica isotérmica (ITC) e por termoforese em microescala (MST) com a finalidade de confrontar os resultados obtidos.

Uma segunda frente do trabalho, sobre a investigação dos mecanismos de regulação de proteínas que possuem tanto o GGDEF quanto o EAL. Para esse propósito foram analisadas sequências em bancos de dados *online* com essa arquitetura. Para análise da interface de dimerização de EALs, foi usada como modelo a estrutura inédita de um exemplar dessa combinação de domínios, provindas de STM3615¹⁶, além de outras estruturas já disponíveis. As análises de resíduos poderiam explicar, por exemplo, porquê o domínio EAL isolado de STM3615 (resíduos 392-668) foi observado como dímeros, independentemente da presença do c-di-GMP.

Os passos seguidos para execução das análises de proteínas com GGDEF e EAL foram:

- Coletar sequências proteicas do banco de dados *online* Uniprot¹⁷⁴ e de um conjunto reduzido do mesmo banco, de sequências selecionadas e classificadas por Galperin e colaboradores.²⁰⁰
- Isolar e alinhar, das sequências obtidas, apenas os domínios de interesse, GGDEF e EAL;
- Desenvolver linhas de código em linguagem Python que auxiliassem nas etapas de filtragem de sequências e classificação de GGDEFs e EALs cataliticamente competentes ou não;
- Análises de acoplamento estatístico com o programa PFstats²⁰¹⁻²⁰² na busca conjuntos de resíduos que estivessem associados a alguma função regulatória de domínios GGDEF e EAL em tandem.
- Análises de alinhamento e estruturais das regiões de dimerização de proteínas EAL, com a finalidade para concluir novos aspectos determinantes para a dimerização desse domínio.

2 MATERIAIS E METODOLOGIAS

Nessa seção serão apresentados os métodos e protocolos dos experimentos incluídos no trabalho. Para uma análise correta dos resultados foi necessária uma boa compreensão das técnicas biofísicas e bioquímicas abordadas no trabalho. Caso o leitor já possua conhecimento sobre as técnicas, as introduções teóricas tomam caráter opcional (apêndice E-J).

2.1 Buscas e análises por bioinformática

O estudo investigou a função de proteínas de membrana através da caracterização de seus domínios isolados e combinações dos mesmos. Isso foi feito devido às dificuldades técnicas intrínsecas da produção heteróloga de proteínas transmembranares, além disso, a delimitação de uma série de proteínas truncadas eleva potencialmente as chances de obter material solúvel. O trabalho com construções contendo domínios isolados, ou uma combinação deles, também auxilia na descrição desses módulos funcionais. Essa abordagem é comum e tem se mostrado muito eficiente em estudos estruturais e funcionais de proteínas modulares, especialmente em outras proteínas detentoras dos domínios GGDEF e/ou EAL, como por exemplo em LapD,^{120,203} FimX,¹⁰⁹ MorA,¹¹⁶ RbdA¹¹⁹ e YfiN.¹¹¹

As sequências de DNA, que codificam as proteínas PA0847 e HsbD^o foram obtidas nos bancos de sequências *online* pseudomonas.com²⁰⁴ e Kegg.²⁰⁵ Para determinar a abrangência do(s) domínio(s) nas sequências, foram utilizados os programas de predição *online* Pfam^{176,206} e SMART.²⁰⁷ Esses dados de abrangência de domínios auxiliaram na determinação das várias partes das proteínas que seriam produzidas (construções). Parâmetros biofísicos foram computados com o servidor *online* ProtParam.²⁰⁸ Eles incluem, dentre outros, o coeficiente de extinção molar ($M^{-1}.cm^{-1}$ ou em $mg^{-1}.mL$) e o ponto isoelétrico (pI).

Com a finalidade de confirmar essas regiões transmembrana e evitar incluí-las nas construções, as sequências foram submetidas aos programas de predição TMHMM²⁰⁹ e Phobius²¹⁰. Isso porque, comumente, é inviável ou pouco efetivo solubilizar porções transmembrana de proteínas, sendo que, para esse finalidade existem protocolos específicos.²¹¹ O programa Phobius também prediz regiões com peptídeo sinal, que também evitou-se incluir nas construções. O programa de predição para estruturas secundárias PSIPRED²¹² e JPred 4²¹³ foram utilizados para ajudar a evitar que a escolha dos limítrofes das construções interrompesse

^o Consulte o apêndice A para lista completa de proteínas da triagem original.

fitas e hélices preditas, o que poderia gerar proteínas instáveis e assim inviáveis para experimentos.

O programa de predição de estrutura terciária I-TASSER²¹⁴ foi útil para obter dados de estruturas já resolvidas e orientar sobre possíveis ligantes. Buscas por sequências similares foram realizadas utilizando o programas *online* jackhmmer do HMMER²¹⁵ e o PSI-BLAST²¹⁶, ambos com parâmetros padrão. Para processamento em larga escala de sequências proteicas foram usados utilitários do programa HMMER 3.1b2. Análises de alinhamento foram feitas utilizando ClustalW com parâmetros padrão e foram processados através do programa Jalview 2.10.3. Perfis de HMM (*hidden Markov models*) foram ilustrados utilizando o servidor *online* skylign.²¹⁷ Árvores filogenéticas foram criadas com o programa MEGA 7.²¹⁸ Alinhamento de estruturas e processamento de imagens das mesmas foram realizadas e editadas com o programa PymolTM 2.0.6 (versão educacional)²¹⁹ e Coot 0.8.6.1.²²⁰ Listas de contatos hidrofóbicos, hidrofílicos, ligações hidrogênio e outros, de estruturas cristalográficas resolvidas, foram obtidas com o programa PIC (*Protein interactions calculator*).²²¹ Todas as imagens tiveram a edição final no programa Inkscape 0.92.²²²

2.2 Obtenção e manipulação dos DNAs codificantes das proteínas

2.2.1 Desenho de primers e plasmídeo utilizado

Os procedimentos para clonagem foram etapas prévias necessárias para produção das amostras. Algumas das bases dessa técnica estão expostas no apêndice E. A etapa inicial é a de desenho de primers, feita a partir das sequências de bases das regiões codificantes de interesse. As submetidas ao programa OligoPerfectTM.²²³ Foram ajustadas, no programa, as faixas de temperatura de desnaturação (T_m – *melting temperature*) em torno de 60°C, uma faixa de porcentagens das bases GC em torno de 50% e um tamanho em torno de 20 bases. O parâmetro maleável adotado foi o T_m dos iniciadores, desde que as opções de primers sense e antisense para esse parâmetro seja a mais próxima possível, com diferença menor que 2°C (convenção arbitrária). Isso porque, com a adição de DMSO ou alterando a concentração de reagentes como $MgCl_2$, pode-se diminuir o T_m dos primers. Caso o T_m ficasse muito grande, maior que 80°C, o tamanho do primer era ajustado, desde que não fique menor que 16 pb (pares de base). Além da sequência que deve se hibridizar ao gene, cada um dos primers deve conter nas extremidades 5' o sítio de restrição escolhido e ainda bases extras para auxiliar no ancoramento da enzima de

restrição. Nesse trabalho, foi padronizado como sequência adicional as bases *cagt*.^P Vetores comerciais apresentam diversos sítios e dois deles devem ser escolhidos de maneira que estejam ausentes região codificante da CDS de interesse, do inglês *coding sequence*. Foi usado para efetuar essa averiguação o programa *online* NEBcutter V2.0²²⁴⁻²²⁵ (New England Biolabs[®]/NEB). As tabelas finais de todos os primers utilizados no trabalho encontram-se no apêndice B.

Foi utilizado para expressar as construções o vetor de expressão pETSUMO, que é o pET28 (Novagem) modificado. Ele foi desenhado para expressão da proteína de interesse ser fusionada com outra denominada SUMO (proteína de fusão, *small ubiquitin-like modifier*). Essa proteína é reconhecida por aumentar o rendimento da purificação por possuir elevada solubilidade.²²⁶ A SUMO de fusão é expressa também com a clássica sequência de seis histidinas (6xHis), necessária para purificação em resina de afinidade.

2.2.2 Protocolo padrão para PCR

As PCRs para obtenção das CDS de interesse são compostas pelos reagentes a seguir:

- DNA molde foi utilizado 200 ng de DNA genômico ou 10 ng de um vetor contendo a região de interesse;
- a polimerase High Fidelity (Fermentas-Thermo Scientific), 1,5 unidades por reação;
- 1mM de dNTPs;
- 5 pM do iniciador sense e 5 pM do antisense;
- água destilada e deionizada em q.s.p 50 µL.

A reação montada segue para termociclador (equipamentos C1000 TouchTM Thermal Cycler ou T100TM Thermal Cycler, (BIO-RAD)). Foi programado nos equipamentos para que as reações fossem submetidas as seguintes etapas, que estão esquematizadas na Tabela 4.

Amostras de 2-3 µL da PCR foram diluídas em tampão *loading buffer* e foram analisadas por eletroforese de agarose 0,8% em tampão TAE e corados com brometo de etídeo (3 µL/100mL de gel). São aplicados um padrão de massas moleculares padrão para comparação. Os padrões utilizados foram 1 Kb PLUSTM DNA Ladder (GOLDBIO) e PageRuler Plus (Thermo Scientific). A fotodocumentação foi feita no equipamento Gel DocTM EZ Imager (BIO-RAD), utilizando o programa Image Lab 4.1 (BIO-RAD). Nos casos em que a banda no

^P Em certos vetores, como pETDuet, é crucial o acréscimo de bases ao primer sense, entre as regiões do sítio de restrição e de hibridização da CDS de interesse. Isso é necessário para corrigir a fase de leitura em relação ao sítio de tradução do ribossomo (RBS) presente no vetor.

gel de agarose se apresentaram com tamanho esperado da CDS, foi procedido com nova eletroforese para extração do gel. O gel contendo o amplificado seguiu para a purificação, feita com o *kit* Wizard® SV Gel (Promega).

Tabela 4 - Esquematisação dos ciclos do termociclador nos quais foram submetidas as PCRs.

Repetição	1X	35X**			1X	1X
Ciclo	Desnaturação inicial	Desnaturação	Hibridização	Polimerização	Polimerização final	Estocagem
Temperatura	95 °C	95 °C	57 °C *	72 °C	72 °C**	4 °C
Tempo	10 min	30 s	30s	1 min / kpb da CDS**	10 min	-

* Pode haver modificação dependendo da temperatura calculada para os primers, fornecida pelo fabricante, ou se é verificada ausência da amplificação em tentativas anteriores.

** Dados fornecido pelo fabricante para polimerase High Fidelity® (Fermentas-Thermo Scientific).

Fonte: Elaborada pelo autor

Nos casos com ausência ou baixa amplificação, o protocolo foi repetido com aditivos ou com alterações nos passos no termociclador. As opções de alterações são listadas a seguir:

- acréscimo de BSA a 0,5 mg/mL.
- adição de 3% (vol/vol) de DMSO;
- uso como molde uma das bandas de uma PCR com ao gene do produto de uma PCR inespecífica;
- variação na temperatura de hibridização como por exemplo, numa tentativa inicial foram utilizados 55°, 61°, 63°, 64° e 65° e volumes de reação de 20 µL. Dependendo dos resultados, novas tentativas eram executadas;

2.2.3 Reação de ligação das CDS em vetor

Para a reação de ligação convencionais, foram montadas, separadamente, reações para o vetor e o inserto com enzimas de restrição, de acordo com aquelas escolhidas no planejamento das clonagens. Para reações foram utilizadas as enzimas FastDigest (Thermo Scientific). Na reação com o plasmídeo foram utilizadas:

- 1 unidade de cada uma das duas enzimas;
- em torno de 1µg de vetor;
- tampão FastDigest (1X);
- água destilada e deionizada, q.s.p. 20 µL de reação.

Para a reação de digestão do produto de PCR purificado são utilizados:

- em torno de 200 ng do DNA;
- 1 unidade de cada uma das duas enzimas;

- tampão FastDigest (Thermo Scientific);
- água destilada e deionizada, q.s.p. 20 µL de reação.

As reações de digestão foram incubadas a 37°C por 12 h à 16 h (*overnight* – abreviado para O.N.). Ao final as reações digestão foram submetidas a separação por gel de agarose padrão 0,8% em tampão TAE e corados com brometo de etídio. As bandas do gel correspondentes ao DNA de interesse são cortadas, tanto para as reações contendo o DNA de interesse quanto para vetor. As amostras em gel foram purificadas usando o *kit* Wizard® SV Gel (Promega).

Para a reação de ligação foram utilizados:
contendo diluído 1 vez e com:

- 50 ng a 150 ng de vetor digerido, se a reação já foi montada uma vez e não foi obtido sucesso;
- inserto numa proporção em tamanho de pares de base de 1:3 (vetor: inserto). Proporções diferentes são usadas em novas reações caso de falha do protocolo (próximo passo de transformação sem colônias);
- 1 unidade da enzima T4 DNA ligase (Fermentas, Thermo Scientific);
- tampão comercial para T4 ligase (Fermentas, Thermo Scientific);
- água destilada deionizada q.s.p. 20 µL.

Para que a ligação ocorra, a reação foi submetida a sequências de passos no termociclador de 4 horas a 22°C e 9h a 16°C, permanecendo armazenado a 4°C após disso. Então 10 µL da reação de ligação são usados para transformação em células de propagação DH5α, seguindo o protocolo de transformação por choque térmico.

Reações alternativas, utilizando o mínimo possível de acréscimo de água foram realizadas. Essas reações não seguiram para o termociclador, mas foram incubadas por apenas 1 h a 1h 30 no gelo. Depois todo o volume de reação seguiu para protocolo de transformação convencional. Essas reações se mostraram bem mais eficientes, ou seja, com maior número de colônias transformantes resultantes (seção 2.2.4).

2.2.4 Protocolo de transformação padrão por choque térmico

O protocolo de transformação é usado para dar prosseguimento ao protocolo de ligação e para obtenção de células de propagação para propagação de plasmídeos fechados no geral (obter mais vetor em solução). Esse protocolo também foi usado para obtenção de células de

expressão com o vetor desejado. Para propagação, foram utilizadas as cepas *E. coli* DH5 α (Novagem), preparadas com protocolo próprio para choque térmico por cloreto de cálcio. As células competentes de expressão foram as cepas de *E. coli* Rosetta ou BL21(DE3) (Novagem).

Iniciou-se o protocolo incubando 50 μ L de células competentes por aproximadamente 15 min no gelo com aproximadamente 50 ng do vetor que deseja transformar. Essa quantidade de DNA foi utilizada para não sobrecarregar a placa com colônias transformantes, impedindo o isolamento destas.

Após isso, a mistura célula e DNA foi incubada por 2 min a 42°C. Rapidamente o tubo com a amostra foi retornada ao gelo, onde permaneceu por mais 5 min. Ao tubo foi adicionado então 150 μ L de meio de cultura LB previamente estéril, Luria Broth (SIGMA-ALDRICH®), ainda sem o antibiótico específico para a seleção do plasmídeo. A mistura seguiu para equipamento de agitação shaker a 37°C e rotação de 250 rpm. A amostra permanece no equipamento por volta de 45 minutos a 1 h 30 min. Após isso, o meio com as células foram plaqueados em meio LB estéril contendo 15% (m/ vol) de ágar (SIGMA-ALDRICH®) e ainda o(s) antibiótico(s) compatível(is) com o vetor e com a célula. Foram utilizados canamicina a 50 μ g/mL para pET29 ou pETSUMO e ampicilina para o vetor pETDuet. Para uso da Rosetta foram adicionados ainda cloranfenicol 35 μ g/mL no meio de cultivo. As placas foram mantidas em estufa a 37°C por média de 12 horas a 16 horas (O.N.), tempo necessário para colônias se tornarem visíveis.

2.2.5 Propagação e confirmação de clones

Iniciou-se o procedimento com a transferência de colônia única de uma placa pré-transformada para uma cultura de 6 mL de LB contendo o antibiótico próprio do plasmídeo e da célula. O pré-inóculo passou então para cultivo em agitador shaker a 37 °C e 250 rpm. Dessa cultura foi extraído o DNA plasmidial, de acordo com o protocolo do *kit Wizard® Plus SV Minipreps DNA Purification System* (Promega). As amostras resultantes seguiram para confirmação com protocolo de análise de restrição. As amostras confirmadas seguiram para o sequenciamento, realizado no laboratório de Biofísica do IFSC-USP, pelo método de Sanger. Nos resultados de sequenciamento foram verificados, por exemplo, se a CDS de interesse está de fato inseridos e no quadro de leitura correto ou se houve mutações ou deleções no DNA codificante da proteína.

2.2.6 Análise de restrição

A reação consiste em 500 ng de DNA com 0,5 unidade de cada uma das enzimas de restrição e o tampão FastDigest (Thermo Scientific). A reação permaneceu em torno de 3 horas a 37°C (em estufa). Após isso, a amostra segue para análise em eletroforese em gel de agarose 0,8%, montado com tampão TAE e contendo brometo de etídio. Se confirmada a presença de um DNA no gel com um tamanho correspondente ao tamanho do gene de interesse, a amostra do vetor obtido na extração segue para sequenciamento.

2.2.7 PCR para mutação sítio-dirigida e para vetores inteiros

Essa seção é referente aos procedimentos de amplificação utilizadas para vetores inteiros, que inclui mutação sítio-dirigida e para vetores do duplo híbrido. Para realizar as mutações sítio-dirigidas, trocas de bases de acordo com o aminoácido que se quer trocar na proteína de interesse se realiza também uma PCR, com algumas diferenças. Como DNA molde foi utilizando o vetor contendo o DNA de interesse, os primers utilizados envolvem a região na qual se quer fazer a mudança e na reação se amplifica todo o vetor. Os primers foram desenhados utilizando a ferramenta *online* QuikChange Primer Design, fornecido gratuitamente pelo site da Agilent (Apêndice B, Tabela 28).²²⁷

A reação de PCR para mutação sítio-dirigida foi composta por:

- 10 ng do vetor com o gene que se deseja mutar;
- 2 pmol/ μ L de cada um dos primers;
- 0,2 mM de uma solução de dNTPs;
- DNA polimerase comercial Pfu (Biolabs) ou Phusion (Thermo Fisher Scientific);
- tampão próprio para as enzimas;
- água destilada e deionizada em q.s.p 50 μ L.

As PCRs de cada um dos genes ocorreram no termociclador, programado para realizar os seguintes ciclos apresentados na Tabela 5. Após a reação, foi submetida a clivagem do DNA molde (metilado) com DpnI (Fermentas). Tal reação foi semelhante a reação de digestão por enzimas de restrição. É necessário adicional 1 μ L de enzima a reação de PCR, homogeneizar, e deixar incubando a 37 °C em estufa ou em um termociclador. Então 10 μ L da reação são avaliados em gel de agarose 0,8 % em tampão TAE. O acréscimo de aditivos, como DMSO (3% no meio reacional) se faz necessária frente a ausência da amplificação ou com amplificações inespecíficas.

Tabela 5 - Resumo esquemático dos ciclos utilizados no termociclador para PCRs de mutação sítio-dirigida.

Repetição	1X	18X			1X	1X
Ciclo	Desnaturação inicial	Desnaturação	Hibridização	Polimerização	Polimerização final	Estocagem
Temperatura	95 °C	95 °C	60 °C *	68 °C	68 °C**	4 °C
Tempo	10 min	30 s	30s	15 s / kpb do vetor com a CDS de interesse **	10 min	-

* Pode haver modificação da temperatura, dependendo daquela calculada para os primers, fornecida pelo fabricante, ou se é verificada ausência da amplificação em tentativas anteriores.

** Dados estipulados pelo fabricante da enzima Pfu® (Fermentas-Thermo Scientific).

Fonte: Elaborada pelo autor.

Quando se obteve resultado positivo no gel, a reação seguiu para transformação em células de propagação DH5α cultivadas o antibiótico adequado ao vetor, como descrito na seção 2.2.6. Quando a placa de cultivo apresenta colônia(s), as mesmas seguem para protocolo de propagação e confirmação de clones.

2.2.8 Clonagem ligase independente para vetores do duplo híbrido

O protocolo de clonagem tem principal característica a não utilização de T4 ligase (LIC, *ligase independent cloning*). Essa técnica foi própria para clonagens em larga escala e foi adotada no trabalho da montagem de bibliotecas de triagens por parceiros de interação de Andrea Cardoso.¹³ A técnica de clonagem por LIC foi tema de trabalhos do Dr. César Camilo²²⁸ que, em colaboração com nosso laboratório, adaptou a técnica para os vetores do duplo híbrido bacteriano *BacterioMatch II Two-Hybrid System*.

Inicialmente, foram realizadas a amplificação não só da CDS das proteínas de interesse, mas também dos vetores pTRG, ou vetor presa, que exige o uso de tetraciclina, 12,5 µg/mL, ou pBT, ou vetor isca, que requer uso de cloranfenicol 34 µg/mL. Para as CDS aplica-se o protocolo de PCR padrão e para amplificação dos vetores foi utilizado o protocolo de PCR de vetores inteiros, semelhante àquela empregada para mutação sítio-dirigida.

Para a técnica, os primers das CDS foram desenhados de maneira convencional (seção 2.2.1) entretanto, a sequência adicional na extremidade 5' é específica para o sistema de clonagem, uma para o primer sense e outra para o antisense. Essas sequências específicas promovem a hibridização entre vetor-inserto e permite a ligação, dispensando o uso de enzimas de restrição. O inserto pode seguir tanto para a ligação no pBT ou no pTRG pois esses vetores permitem foram alterados para fornecer a mesma região de hibridização. Para amplificação dos

vetores foram utilizados oligos próprios pois foram originalmente planejados para manipulação com enzimas de restrição. Procede-se com protocolo de PCR padrão. Como trata-se de sequências grandes deve-se utilizar tempo de extensão maior para a amplificação dos plasmídeos e recomenda-se fortemente o uso de uma enzima de alta fidelidade, Pfu (Biolabs) e Phusion (Thermo Fisher Scientific).

Com os amplificadores purificados prosseguiu-se com o tratamento com T4 polimerase, tanto do inserto quanto de cada um dos plasmídeos. O protocolo foi adaptado²²⁸ e foi utilizado para criar as extremidades coesivas para hibridização entre inserto e vetor. A reação foi composta como segue, válido para enzima da T4 polimerase (Thermo Scientific):

- DNA molde, 500 ng para a reação com o vetor e 300 ng para reação com o inserto codificante;
- 3 unidades de T4 polimerase (Thermo Scientific);
- 2 µL de dXTP a 10 mM, dTTP se for a reação com um dos vetores e dATP para o inserto codificante;
- Tampão 1X (Thermo Scientific);
- 1 µL DTT 100mM;
- água destilada e deionizada em q.s.p 20 µL.

O meio reacional permaneceu por 30 minutos a 25°C. A enzima foi inativada, submetendo a reação por 20 min a 80°C. A seguir, 2 µL da reação montada para o vetor (50ng) foi misturado com uma quantidade de inserto necessária para a proporção 1:3 (vetor: inserto) e foi completado o volume para 10 µL. Foi adicionado então a essa mistura 40µL do tampão de transformação gelado, composto por 100 mM KCl, 30 mM CaCl₂, 50 mM MgCl₂ e 1,5 % (m/vol) PEG 4000. Essa mistura final permanece 15 minutos no gelo para então ser utilizada para transformação por choque térmico. O protocolo recomenda que se use células de propagação próprias *XLI-Blue MRF'*, fornecida pelo fabricante (Agilent) pois são tóxicos, e ainda recomenda fazer o crescimento a 30°C.

As colônias resultantes eram usadas no protocolo de extração plasmidial com o uso do *kit Wizard® Plus SV Minipreps DNA Purification System* (Promega). Para a confirmação da inserção do DNA deve-se realizar uma PCR convencional. Nessa, era utilizada ainda uma polimerase de baixa fidelidade como a Taq Polimerase (Thermo Scientific), já que o objetivo era verificar o tamanho do inserto. Eram utilizados para essa PCR, os primers para amplificação das CDS ou para o sequenciamento dos vetores (apêndice B). A segunda opção foi mais aconselhada quando se tem vários clones diferentes.

2.3 Obtenção das proteínas

2.3.1 Teste de expressão e purificação

Uma vez obtido o vetor contendo o inserto codificante de interesse, foi realizado o protocolo de teste de expressão junto com uma purificação em baixa escala. Nesse teste, foram recolhidas amostras de todos os passos da purificação para verificar se alguma, ou algumas, das etapas do protocolo eram inadequadas para a proteína de interesse.

O tampão inicial padrão assumido para purificação foi:

- 25 mM tris;
- 300mM NaCl;
- pH 8,0;

Era observado se o pH utilizado não coincidia com o pI da proteína, ou se ficava bem próxima ($< 0,5$), o que poderia impedir o prosseguimento com o tampão padrão e indicava a necessidade da troca do mesmo. Dada essa verificação, para expressão era feito o preparo de um pré-inóculo de 6 mL de LB, a partir de colônia única, de uma placa transformada previamente com protocolo convencional e os antibióticos indicados (seção 2.2.4). Esse pré-inóculo foi crescido a 37°C sob agitação de 250 rpm, por aproximadamente 16 horas (*overnight*, abreviado para O.N.). Foram usadas como padrão células BL21(DE3) (Novagem). Frente a resultados negativos, foi testado o mesmo protocolo com Rosetta (Novagem). Do pré-inóculo pronto, 4 mL foram usados no inóculo de LB estéril de 200 mL contendo os antibióticos. A proporção, então, de pré-inóculo e inóculo foi de aproximadamente 1:50. O inóculo fica sob agitação por 37°C a 250 rpm até a densidade óptica (D.O) atingir um valor entre 0,6 e 0,8. Atingida essa DO, foi acrescido ao meio quantidade suficiente de IPTG para uma concentração final de 0,5 mM. Então, cultivo permaneceu no agitador para expressão, por mais 3h. Frente a resultados negativos, poderia ser trocada a temperatura de expressão para 18 °C, a quantidade de IPTG reduzida para 0,1 mM e o tempo de indução para 16 h. Ao final da expressão, uma alíquota de 100 µL foi separada para amostra para análise em gel de poliacrilamida. A cultura final foi centrifugada a 5 000 g por 20 min. e o meio foi descartado. As células foram resuspendidas em 10 mL do tampão de purificação contendo 10 mM de imidazol adicionais.

Seguiu-se com a lise das células, na qual primeiro foi feita a incubação da solução com as células por 20 min no gelo com 2 mg de lisozima (10mg para cada litro de cultura). Depois foi realizada a sonicação das células com o equipamento Sonic Dismembrator Model 500 (Fisher Scientific), configurado com 15 % amplitude em funcionamento por 5 min, com

intervalos de 30 s ligado e 30 s desligado. Por fim, o sobrenadante das células era separado, usando centrifugação de 30 min a 12 000 g, filtrado e aplicado a resina de cobalto. A quantidade de resina utilizada no experimento foi de 200 μ L (produto Talon, fabricante Clontech). Previamente foi necessário passar 20 VC (volumes de solução/volume de resina de afinidade) de água e 20 VC de tampão na resina, etapa de equilíbrio. A proteína de fusão no experimento se liga a resina, enquanto outros componentes provindos da célula de expressão, na maioria, passam direto pela coluna. Essa amostra é o *flow through* (FT), ou não ligado (NL). Foram aplicados a seguir 10 VC de tampão de purificação contendo 5 mM ou 10 mM de imidazol para remoção de ligantes inespecíficos. Depois, aplicou-se mais 10 VC de tampão de purificação sem imidazol (etapa de lavagem). Ao final foi feita a eluição com o tampão padrão contendo 500 mM de imidazol. A proteína foi clivada por 16 horas a 8°C com excesso de 0,2 mg SUMO protease (ou Ulp, abreviação da proteína de origem *ubiquitin-like-specific protease 1*). Isso equivale a uma proporção de, aproximadamente, 1:100 (protease: proteína de fusão). A amostra em seguida foi submetida a uma etapa de dessalinização, realizada no cromatógrafo com a coluna HiPrep 26/10 Desalting (GE Healthcare) para trocar o tampão para o padrão, sem imidazol. A amostra passou então por uma segunda cromatografia de afinidade para remoção da proteína de fusão SUMO e da protease Ulp. Etapas de lavagem adicionais eram realizadas, com a aplicação de 2 VC de tampão de purificação sem imidazol, (primeira lavagem, L1), seguida de 8 VC do mesmo tampão (segunda lavagem) e por último era procedido com a eluição do SUMO, aplicando na resina 10 VC tampão de purificação contendo 500 mM de imidazol.

Caso houvesse amostra suficiente, a mesma era concentrada e submetida a uma exclusão molecular. Para isso, previamente as amostras foram concentradas em sistemas de concentrador em tubo de 50 mL (Milipore). As colunas utilizadas foram Superdex 75 10/300 GL e Superdex 200 10/300 GL (GE Healthcare) acoplada a um cromatógrafo (AKTA pure, GE Healthcare). Para a exclusão molecular deveria haver 1 mL de amostra, com aproximadamente 0,5 absorbância no 280nm. Isso era feito para se obter um sinal bom no cromatograma, com por volta de 50 mUA (unidades arbitrárias de absorbância) para o pico da eluição da proteína (diluição de 10 vezes, aproximadamente), caso haja uma única população (oligomérica ou conformacional).

Amostras dos passos da purificação foram coletadas e analisadas em gel SDS-PAGE 15%, de acordo com o método Laemmli.²²⁹ A análise foi feita para verificar se os passos da purificação foram bem-sucedidos, avaliar o grau de pureza das amostras e certificar se as massas observadas estão próximas do esperado. Para estimar as massas foi realizada a

comparação com um padrão de massas moleculares (MM), também aplicado junto com as amostras no gel. Foram utilizados os padrões BLUEstain™ 2 Protein Ladder (Gold Biotechnology Inc). e outros preparados no laboratório de biofísica do IFSC-USP. São recolhidas 40 μ L das amostras dos passos da purificação, que são do sobrenadante (S) da centrifugação para remoção dos debris celulares, não ligado (NL), lavagens com tampão, amostra eluída (500mM), amostra eluída clivada (CV) com Ulp, lavagens da segunda afinidade e eluição da segunda afinidade. A cada uma dessas alíquotas foi adicionado 20 μ L de tampão 3X concentrado de tampão de amostra com beta-mercaptoetanol. As amostras então eram fervidas por 10 min em banho seco a 95 °C, etapas usadas para desnaturação.

Figura 36 - Esquemática simplificada dos passos de purificação com protocolo padrão.

O primeiro passo foi inserir o plasmídeo em uma bactéria própria para expressão. Essa bactéria foi cultivada e expressa a proteína com a adição de IPTG ao meio. As células expressas são separadas do meio para lise celular e liberação das proteínas. Foi utilizada a expressão com a fusão SUMO, que possui na região N-terminal seis histidinas (6xHis-Tag). Essas histidinas permitem a purificação por apresentar afinidade pela resina. Posteriormente a porção 6xHis e SUMO foi clivado pela SUMO protease (Ulp). Etapas de cromatografia de exclusão visaram elevar a pureza e permitir análises de população da proteína de interesse.

Fonte: Elaborada pelo autor.

2.3.2 Protocolo de purificação e etapas de otimização

O protocolo empregado para purificação seguiu alguns dos passos do teste, mas com algumas alterações, de acordo com as particularidades de cada proteína. O tampão padrão foi o mesmo do teste, 25 mM tris, 300mM NaCl em pH 8,0. Entretanto, dependendo do experimento em que iria ser utilizada a proteína, o tampão passa para 25 mM de hepes ou outro segundo passo de otimização do tampão. Por exemplo, em experimentos baseados em grupos reativos com amina, o tris não é adequado, por possuir esses grupos. Concentrações salinas de 150mM e 600 mM também foram utilizadas para observar mudanças em populações oligoméricas, na atividade e verificar diminuições na polidispersividade das amostras, medidas no equipamento de espalhamento de luz dinâmico (DLS).

De maneira análoga aos testes, a expressão utilizou um pré-inóculo de 6 mL de LB, preparado com colônia da placa transformada previamente pelo protocolo convencional e os antibióticos indicados. Esse pré-inóculo foi crescido a 37°C sob agitação de 250 rpm por aproximadamente 16 horas (O.N.). Inóculos de 1 litro foram montados em erlenmeyers de 2 litros de maneira estéril, contendo todo o conteúdo do pré-inóculo e antibiótico da célula e do vetor. Nos casos em que as purificações forneceram baixo rendimento por litro de cultura, mais de um litro de cultivo era utilizado para purificação. O crescimento celular foi feito por agitação a 37°C e 250 rpm, até a densidade óptica (D.O) atingir um valor entre 0,6 e 0,8. Isso levava em torno de 2 h 30 a 3 h. As células foram induzidas dessa vez com 0,1 mM de IPTG e a expressão foi realizada a 18 °C e 250 rpm. O meio resultante foi centrifugado e as células resuspendidas em 20 mL do tampão de purificação contendo 5 mM ou 10 mM de imidazol.

A lise das células era iniciada incubando-se em gelo as células com 10 mg de lisozima para cada litro de cultura. A sonicação foi feita no equipamento Sonic Dismembrator Model 500 (Fisher Scientific) com 15 % amplitude por 7-10 min, com intervalos no equipamento de 30 s ligado e 30 s desligado. Uma opção para realizar a lise de muito volume de células, mais daquele resultante de mais de 2 litros de cultura, foi realizar o procedimento no sonicador de ponta espessa, no equipamento Fisherbrand™ Q500 Sonicator (Fisher Scientific), configurado com 35% amplitude, 30 s desligado e 59 segundos ligado, com tempo total ligado de 7 – 10 min. Por fim, o sobrenadante foi separado usando centrifugação por 30 min a 12 000 g, filtrado com papel filtro Qualy (J Prolab) e aplicado a resina de afinidade, previamente equilibrada com 20 VC de tampão. Foram utilizados 2 mL de resina de cobalto (produto Talon, fabricante Clontech) para cada litro de cultura. Após aplicar a amostra do sobrenadante na resina, era realizada a etapa de lavagem com 10 VC de tampão de purificação contendo 5 mM ou 10 mM de imidazol

para remoção de ligantes inespecíficos. Depois aplica-se mais 10 VC de tampão de purificação sem imidazol.

Nesse passo da purificação foram incluídas etapas que envolviam a incubação com a fosfodiesterase STM3615²⁴²⁻⁶⁶⁸ para proteínas com domínio GGDEF e que foi verificada a copurificação com o c-di-GMP. A STM3615²⁴²⁻⁶⁶⁸ estava disponível no nosso laboratório e sua atividade já havia sido verificada no trabalho de doutorado de Flávio Rosseto¹⁶ e por isso foi possível utilizá-la. Essas etapas adicionais serviram para remover o dinucleotídeo copurificado com as enzimas. Entretanto, foi necessário verificar previamente se STM3615²⁴²⁻⁶⁶⁸ utilizada era ativa, como descrito a seguir (seção 2.4.3).

Uma das opções de tratamento era adicionar, na própria resina contendo ligada a proteína em fusão, 10 mL do tampão de purificação, uma quantidade de PDE suficiente para aproximadamente 0,7 μ M e 2 mM $MgCl_2$ (concentrações finais). A mistura era gentilmente homogeneizada e era reservada, em temperatura ambiente ($\sim 22^\circ C$), por 3 h. Após isso, o tampão adicionado era descartado e então era realizada uma etapa de lavagem, com 20 VC com tampão padrão. Caso, no purificado final, fornecesse sinal de c-di-GMP em análise por HPLC, um segundo procedimento era realizado. Nele, a amostra era eluída com 5 VC de tampão de purificação contendo 500 mM de imidazol. Essa amostra passava por procedimento de dessalinização, realizada no cromatógrafo com a coluna HiPrep 26/10 Desalting (GE Healthcare) para remoção do imidazol. Então, era adicionado em torno de 1,4 μ M de PDE e 2 mM de $MgCl_2$ a solução contendo a proteína em fusão com 6xHis-SUMO. A solução era reservada na geladeira ($4-8^\circ C$) por 16 h. No dia seguinte a amostra submetida a nova purificação por cromatografia de afinidade, com resina equilibrada com tampão sem imidazol. Nessa etapa a PDE não interage e foi eluída. Uma lavagem com 20 VC com tampão padrão era realizada.

O próximo passo consistia em clivar a fusão, na própria resina de afinidade, aplicando 10 VC de tampão contendo 0,2 mg de Ulp para cada litro de cultura. A resina era gentilmente homogeneizada e reservada na geladeira ($4-8^\circ C$) por aproximadamente 16 h. Após a clivagem a amostra o tampão era recolhido e uma etapa de lavagem com 2 VC era realizado. Soluções as quais detinham as amostras. Uma etapa de eluição era feita com tampão padrão contendo 500 mM de imidazol, devendo conter Ulp e a SUMO.

As proteínas previamente foram concentradas em sistemas de concentrador em tubo de 50 mL (Milipore). concentrada submetida a cromatografia de exclusão molecular, com as colunas Superdex 75 10/300 ou Superdex 200 10/300, (GE Healthcare). Amostras finais então eram analisadas em gel de SDS-PAGE 15 % para verificar se os a pureza nas amostras e

certificar se as massas observadas estão próximas do esperado, através da comparação com padrão comercial BLUEstain™ 2 Protein Ladder (Gold Biotechnology, Inc) ou de confecção própria, do laboratório de biofísica do IFSC-USP. Eram recolhidas 40 µL de cada amostra que se desejava analisar, que eram misturadas com 20 µL de tampão 3X concentrado de amostra com beta-mercaptoetanol. Essas eram fervidas por 10 min em banho seco a 95°C. Essas amostras então seguiam para análise em gel SDS-PAGE 15 %.

2.4 Produção de c-di-GMP e análises por cromatografia de fase reversa

2.4.1 Produção de c-di-GMP e purificação de c-di-GMP por cromatografia de fase reversa

O protocolo foi adaptado daqueles apresentados em trabalhos de Zahringer⁸⁵ e Navarro.¹¹³ Foi utilizada como enzima, a princípio, a enzima WspR, produzida e purificada em nosso laboratório. Após a confirmação de que a construção PA0847-HPG-3 (resíduos 369-678) produz c-di-GMP (seção 2.4.4), essa construção passou a ser utilizada no protocolo. O meio reacional era constituído por:

- 5 µM de enzima,
- 500 µM de GTP, previamente diluído e quantificado em água ($\epsilon_{\text{Molar}} = 26\,100\text{ M}^{-1}\cdot\text{cm}^{-1}$);
- tampão composto por 5 mM MgCl₂, 50 mM tris e 50mM de NaCl em pH 8,0.

As reações foram realizadas em tubos falcon de 50 mL. A reação foi reservada sob suaves agitações em plataforma oscilante por aproximadamente 16 h (O.N.). Dado esse tempo, os tubos foram submetidos a banho quente a 98°C por 10 min. Logo após o meio reacional foi esfriado a temperatura ambiente (~22 °C), separado em tubos de 2 mL e centrifugado por 20 min a 16 000 g, velocidade máxima da centrífuga de bancada. O sobrenadante foi então filtrado, com filtro de 0,25 µm. Esse procedimento foi necessário para extração da enzima precipitada remanescente. O produto do processo foi concentrado 10 vezes, utilizando um sistema liofilizador Savant speedvac® PLUS SC210A (Thermo Quest) acoplado a bomba de vácuo.

A purificação foi procedida no cromatógrafo Akta (GE Healthcare) com a coluna de fase reversa preparativa C18 250 x 10 mm código 218TP1010 (VYDAC™) ou Gemini 5 µm 110 Å 250 x 10 mm, código 00G-4435-N0 (Phenomenex® Inc.). A coluna foi equilibrada com o tampão 100 mM acetato de amônio em pH 6,0 (solução A). Foram aplicados 1 mL do concentrado reacional, previamente centrifugados por 10 min a 16 000 g. O c-di-GMP, pela característica apolar, se liga a resina. Então o analito foi eluído com a aplicação de metanol 100 % (solução B), de acordo com as concentrações descritas na Tabela 6.

Tabela 6 - Aplicação dos tampões na coluna de fase reversa para purificação de c-di-GMP.

~ 0 - 32 mL (2,2 VC)	~ 32 - 42 mL (1,5 VC)	~ 42 - 61 mL (1,5 VC)	~ 61 - 62,5 mL (0,1 VC)	~ 62,5 - 86 mL (0,6 VC)	~ 86 - 87,5 mL (0,1 VC)	~ 87,5 - 120 mL (2,2 VC)
0 % B	Gradiente 0 - 30 % B	30 % B	Gradiente 30 - 90 % B	90 % B	Gradiente 90 - 0 % B	0 % B

Fonte: Elaborada pelo autor.

Os picos foram analisados por espectrometria de massas. A fração correspondente a eluição do c-di-GMP, destacado no cromatograma, foi liofilizado no mesmo sistema do equipamento Savant speedvac® PLUS SC210A (Thermo Quest). O rendimento foi em torno de 20 % de conversão GTP para c-di-GMP. O fluxo utilizado foi de 1,4 mL/min para as etapas sem metanol e de 1 mL/min nas etapas contendo metanol, decorrente do aumento considerável da pressão, que deveria ser menos que 6 MPa (limite de pressão de trabalho das colunas).

Figura 37 - Exemplificação do cromatogramas de purificação do c-di-GMP.

Perfil de eluição de amostra reacional de produção de c-di-GMP utilizando a coluna da Phenomenex. O pico correspondente ao c-di-GMP esta destacado em amarelo. O tracejado indica o gradiente aplicado durante o experimento.

Fonte: Elaborada pelo autor.

2.4.2 Análises de c-di-GMP e pGpG por cromatografia de fase reversa

Para verificar se o c-di-GMP era copurificado com as proteínas, amostras preparadas das mesmas foram submetidas a experimento de cromatografia de fase reversa. Os métodos de separação são baseados naqueles aplicados para WspR, YdeH (DgcZ) e outras enzimas DGCs com estudos presentes na literatura.^{84,103,113,230}

Para cada amostra proteica purificada que se desejou analisar, foram coletadas 200 µL. Nessa amostra foram adicionados 100 µL de HClO₄ gelado. A amostra foi reservada por 10 min no gelo e segue para centrifugação 10 min a velocidade máxima de centrífuga de bancada de 16 000 g. O sobrenadante dessa amostra foi recolhido para análise. O procedimento foi

adaptado do protocolo apresentado no trabalho de doutorado da Dr^a Joci Macêdo, pois foi julgado interessante se as amostra pudessem também conter GTP ou GDP.²³¹ O sobrenadante seguia então para análise em HPLC, como apresentado na Tabela 7.

A aplicação das amostras e dos diferentes solventes a coluna foi feita com o equipamento de HPLC Waters 2695 Separations Module. Foram aplicadas, sobre fluxo constante de 0,4 mL/min, diferentes composições da solução hidrofílica (A, 100 mM acetato de amônio em pH 6,0) e hidrofóbica (B, metanol). Foram padronizados injeções de 100 µL das amostras e o uso da coluna Gemini 3 µm C18 110 Å 150 x 4,6 mm, código 00F-4439-E0 (Phenomenex® Inc.).⁹

Tabela 7 - Aplicação dos tampões na coluna de fase reversa para análise de nucleotídeos.

0 - 6 min	6 - 7 min.	7- 14 min.	14 - 15 min.	15 – 19 min.	19 – 20 min.	20 - 24 min.
0 % B	Gradiente 0 - 30 % B	30 % B	Gradiente 30 - 100 % B	100 % B	Gradiente 30 - 100 % B	0 % B

Fonte: Elaborada pelo autor.

2.4.3 Cinética fosfodiesterase com EAL

Esse experimento foi utilizado para dimensionar a eficiência catalítica da construção citoplasmática da proteína STM3615 que abrange os domínios GGDEF e EAL, abreviado para STM3615^{GGDEF-EAL} ou STM3615²⁴²⁻⁶⁶⁸. Isso foi feito para verificar sua viabilidade para uso para remoção do c-di-GMP copurificado em construções que possuem o domínio GGDEF. O procedimento foi baseado naquele apresentado no trabalho de RocR.⁸³ O meio reacional utilizado foi montado, contendo:

- 100 mM de tris pH 8,0;
- 20 mM de KCl;
- 2 mM de MgCl₂;
- 1 µM da STM3615²⁴²⁻⁶⁶⁸;
- 32 µM de c-di-GMP (a adição do dinucleotídeo marca o início da contagem do tempo de reação).

Em alíquotas de 140 µL de reação foram adicionadas 14 µL de CaCl₂ (proporção 1:10) para inibir a atividade PDE e parar a reação. As amostras então foram fervidas a 95 °C,

⁹ Outros trabalhos utilizam para eluição bicarbonato de trietilamônio ou outras colunas de fase reversa menores, como a Eclipse XDB-C18, dimensões 4,6 x 150 mm (Agilent) ou Atlantis dc18, 4,6 x 50 mm (Waters). Esses trabalhos mostraram melhores resultados de separação entre pGpG e c-di-GMP.^{83,232}

centrifugadas a 16 000 g por 10 min e o sobrenadante foi analisado por HPLC, pelo mesmo procedimento utilizado na seção anterior (seção 2.4.2).

Figura 38 - Gradiente utilizado nas análises de c-di-GMP (A) e a resultado da análise de atividade de STM3615²⁴²⁻⁶⁶⁸ utilizando o método apresentado(B).

Fonte: Elaborada pelo autor.

2.4.4 Pré-testes de cinética com análise por fase reversa

O primeiro passo empregado para analisar as proteínas contendo o domínio GGDEF foi constatar qualitativamente se as mesmas eram capazes de converter GTP em c-di-GMP, mesmo que de maneira residual. Para isso, um meio reacional foi montado com um excesso de proteína de 10 μ M ou 100 μ M. Essa reação possui 2 mL e era constituída do próprio tampão de purificação (tris-HCl 25mM, NaCl 300mM em pH 8,0) ao qual foi adicionado 5 mM do cofator $MgCl_2$ e 500 μ M do substrato GTP. Aliquotas de 200 μ L foram separadas nos tempos de 1, 3 e 10 horas. Cada alíquota passa por inativação da enzima, feita por aquecimento a 95°C por 10 min em banho seco. Centrifuga-se cada amostra a 16 000 g durante 10 min para separação do sobrenadante, que contém o produto da reação.

Foi padronizado injeções de 100 μ L das amostras submetidas a cromatografia em fase reversa utilizando a coluna C18 Nova Pak[®] C18 (dimensões 2,1x150mm). A aplicação dos diferentes solventes a coluna foi feita com equipamento de HPLC (*high-performance liquid chromatography*) Waters 2695 Separations Module. Foram aplicadas, sobre fluxo constante de 0,4 mL/min, diferentes composições da solução hidrofílica (A) e hidrofóbica (B) para eluição dos componentes reacionais. A solução tamponante A é composta por 100 mM de acetato de amônio em pH 6,0 e o solvente B é essencialmente metanol. No método de separação foram aplicadas diferentes composições dessas soluções, com exposto na Tabela 8.

Tabela 8 - Aplicação dos tampões na coluna de fase reversa para análise de atividade DGC.

0 - 6 min	6 - 7 min.	7- 12 min.	12 - 13 min.	13 – 17 min.
0 % B	Gradiente 0 - 70 % B	70 % B	Gradiente 70 - 0 % B	0 % B

Fonte: Elaborada pelo autor.

A identificação dos eluídos foi feita pelo monitoramento nos comprimentos de onda 280 nm e 254 nm da fase móvel eluída através do monitor Waters 2487 HPLC UV-Visible Detector. Os picos de absorbância obtidos foram comparados com o tempo de eluição de amostras padrão de c-di-GMP, provindo de purificações do nosso laboratório (seções 2.4.1 e 2.5.1).

2.5 Comprovação das biomoléculas por espectrometria de massas

Foram realizadas medidas de espectrometria de massas para validação das biomoléculas envolvida no estudo - do c-di-GMP e da fração de HsbD-Nt. As medidas foram feitas com o equipamento micrOTOF-QII (Bruker Daltonics) operado, em colaboração com Dr. Edson Crusca e Dr. Rafael Panhota (IFSC-USP).

2.5.1 Medidas do c-di-GMP por espectrometria de massas

As reações para produção de c-di-GMP, usando uma construção de PA0847 (HPG-3), foram submetidas a experimento de cromatografia de fase reversa C18 preparativa (seção 2.4.1). As frações dessa cromatografia separadas desse experimento seguiram diretamente para análise por espectrometria de massas (Figura 39).

Figura 39 - Resultados de massas das amostras purificadas de c-di-GMP.

As amostras analisadas são dos picos de eluição I (espectro A) e II (espectro B) das purificações, como mostrado na Figura 37. A massa molecular do c-di-GMP é de 690,09 Da. Com o equipamento trabalhando no modo positivo (protonação), era esperado uma massa de aproximadamente 1 Da maior, como foi observado. U.A.: unidades arbitrárias (do equipamento).

Fonte: Elaborada em colaboração com: Dr. Edson Crusca (A); Dr. Rafael Panhota (B).

O tampão da eluição da cromatografia foi compatível com o experimento de espectrometria de massas (acetato de amônio e metanol). O experimento foi feito para garantir que a amostra obtida de fato se tratava do c-di-GMP.

2.5.2 Medidas de amostras proteicas não fragmentadas

As amostras foram submetidas a uma cromatografia na coluna Superdex Peptide 10/300 GL (GE Healthcare) para troca do tampão de purificação para solução adequada para espectrometria de massas. O tampão utilizado foi acetato de amônio 10 mM. A concentração final de proteína para realização de medidas foi da ordem de 10 μ M.

2.5.3 Preparo de amostras proteicas pelo método de tripsinização

Este protocolo foi adaptado pelo CeBiME (Centro de Biologia Molecular Estrutural-CNPq, Campinas), baseado no trabalho de Shevchenko e colaboradores.²³³ Previamente, foi necessário verificar se a sequência da amostra em análise possui sítios para tripsina. Isso foi feito utilizando o programa PeptideCutter do servidor *online* Expasy.²³⁴ Para os procedimentos, foram utilizadas as soluções descritas a seguir (Tabela 9).

Tabela 9 - Soluções utilizadas para extração de gel e tripsinização para análise em espectrômetro de massas.

Solução 1	Metanol 50 % e ácido acético 2,5 % em água destilada e deionizada
Solução 2	Bicarbonato de amônio 100 mM em água destilada e deionizada
Solução 3	Bicarbonato de amônio 50 mM em água destilada e deionizada
Solução 4	Acetonitrila 100 %
Solução 5	DTT 10 mM em solução 2
Solução 6	Iodoacetamina (IAA) 50 mM em solução 2
Solução 7	Tripsina 20 ng/ μ L em solução 3
Solução 8	Ácido fórmico 5 % (vol/vol) em água destilada e deionizada
Solução 9	Ácido fórmico 5 % (vol/vol) e acetonitrila 50 % em água destilada e deionizada

Fonte: Adaptada pelo autor com material fornecido pelo laboratório de Biofísica Sérgio Mascarenhas.

O primeiro passo foi recortar a banda que se quer analisar do SDS-PAGE. Essa banda seguiu para descoloração e remoção do gel com 0,5 mL da solução 1 em microtubo de 1,5 mL. A mistura foi reservada por 1 h a temperatura ambiente (~ 22 °C). A solução foi removida e repetiu-se o procedimento de acréscimo de 0,5 mL da solução 1 e incubação por 1 h a temperatura ambiente.

Para desidratação do gel, foi acrescentado ao microtubo com a banda 200 μL da solução 4 e, após 5 min, a solução foi descartada. Essa etapa foi repetida. Ao final, para remover toda acetonitrila, o conteúdo passou por liofilização com sistema Savant speedvac® PLUS SC210A (Thermo Quest) acoplado a bomba de vácuo, por 2-3 min.

No processo de redução, foi adicionado 30 μL da solução 5 (contendo DTT), ou volume necessário para cobrir o gel recortado. Então segue incubação por 30 min a temperatura ambiente ($\sim 22^\circ\text{C}$). Utilizou-se uma centrifugação rápida (spin) em centrífuga de bancada para remoção da solução 5. Para alquilação, foi feito o mesmo procedimento anterior realizado com a solução 5, mas utilizando a solução 6 (contendo IAA). Em seguida foi realizada a lavagem com a adição de 100 μL da solução 2, com a remoção da solução após incubação por 10 min. Uma nova etapa de desidratação do gel foi realizada, com 200 μL da solução 4 por 5 min, seguida da remoção da solução. A seguir foi feita a reidratação da banda do gel, com adição de 200 μL de solução 2, reservando a mistura por 10 min. Após um spin a solução foi removida para então adicionar 100 μL da solução 4, que seguiu incubação por 5 min. Dado o tempo de incubação, repetiu-se a incubação com solução 4. Ao final foi removido a acetonitrila e o remanescente foi retirado com liofilização.

Foi adicionado então de 30-50 μL de solução de tripsina (7), e a mistura foi reservada em gelo por 30 min. O microtubo foi submetido a um spin para remoção da solução 7. Após isso, foram acrescentados de 5 a 20 μL de solução 3 e a mistura seguiu para incubação a 37°C por 16 h. Dado esse tempo, a solução não foi removida, apenas foram acrescentados 10 μL da solução de extração 1. A amostra foi incubada por 10 min a temperatura ambiente ($\sim 22^\circ\text{C}$). Após isso, foram acrescentados de 10-30 μL , dependendo do tamanho do gel, da solução 2. Segue-se uma nova incubação, por 10 min a temperatura ambiente. Com um *spin*, a solução resultante foi recolhida em um outro microtubo de 1,5 mL. A amostra foi liofilizada até ficar com aproximadamente 50 μL . Então a amostra seguiu para medida no massas.

2.6 Análise de oligomerização

2.6.1 Análises comparativas com padrões por cromatografia de exclusão molecular (SEC) analítica

Para inferir o estado oligomérico, os fragmentos purificados foram analisados por cromatografia de exclusão molecular, abreviado para SEC, do inglês *size exclusion chromatography*. Os procedimentos foram baseados no manual do *kit* comercial *Gel Filtration*

Calibration (GE Helthcare). Como padrão, foi obtido o volume de eluição (V_e) de misturas de proteínas conhecidas que são aprotinina (6,5 kDa), ribonuclease A (13,7 kDa), anidrase carbônica (29,0 kDa), ovalbumina (43,0 kDa), conalbumina (75,0 kDa), aldolase (158,0 kDa) e ferritina (440,0 kDa). Foram preparadas amostras concentradas desses padrões em tampão tris 25 mM, NaCl 300 mM em pH 8,0. Foram diluídas 4 das amostras dos padrões em volume final de 100 μ L em diferentes combinações. Cada padrão foi diluído com equivalente a 1,0 U.A. (280 nm) no volume final. Com essas amostras dos padrões foram realizadas cromatografias de exclusão molecular analítica, utilizando as colunas Superdex 75 10/300 ou Superdex 200 10/300, (GE Healthcare). O tampão utilizado no experimento foi o padrão, 25 mM tris pH 8,0 com 300 mM NaCl. As cromatografias bombeadas e monitoras com o equipamento AKTA pure (GE Healthcare).

Além disso, foi feito uma exclusão molecular apenas com o polímero Blue Dextran 2000 (100 μ L, 3 unidades de absorbância no 280 nm), para obtenção do volume morto da coluna (V_o). Com os valores dos volumes de eluição (V_e), o volume da coluna (V_t) e o volume morto são calculados os coeficientes de partição (K_{av}) de cada amostra através da equação (1). Os dados dos volumes de eluição foram fornecidos pelo próprio método de integração do programa do cromatógrafo UNICORN 6.1 (GE Healthcare).

$$K_{av} = \frac{V_e - V_o}{V_t - V_o} \quad (1)$$

Foram plotados os logaritmos das massas dos padrões contra os coeficientes de partição respectivos. Obtém-se então uma equação de uma correlaciona K_{av} com o logaritmo das massas moleculares para a coluna utilizada. partindo dos volumes de eluição das mesmas:

$$K_{av} = A \times \log(MM) + B \quad (2)$$

Com as amostras em estudo foram realizadas cromatografias de maneira idêntica a realizada com os padrões. Os volumes de eluição das amostras foram convertidos em massa molecular através da regressão apresentada na equação (2). As médias com os padrões e com as amostras foram realizadas na mesma coluna e na mesma batelada de medidas. Isso porque ocorre alteração no perfil de eluição conforme o uso da coluna de exclusão, o gera desvios nos valores calculados.

No cálculo de massas moleculares por SEC são assumidas características que, por vezes, não são válidas para proteínas em estudo. Dentre elas, que as proteínas são globulares, como o padrão utilizado, e que não ocorre interação entre proteína e a resina utilizada para separação. Portanto, desvios da massa esperada podem ser intrínsecos do experimento.

2.6.2 Análises por exclusão molecular acoplada a espalhamento de luz (SEC-MALS)

Análises de espalhamento de luz a múltiplos ângulos acoplada a exclusão molecular, SEC-MALS (*size-exclusion cromathography and multiple angle light scattering*) permitem a obtenção de parâmetros biofísicos de maneira independente da forma do analito em solução e do volume de eluição do mesmo. A montagem experimental esta ilustrada a seguir (Figura 40).

Figura 40 – Montagem experimental do SEC-MALS.

Sobre fluxo constante de tampão, a amostra é aplicada e separada na exclusão molecular. Após a separação, as amostras passam pelo detector de variação de índice de refração (dRI). A amostra segue então para os três detectores de espalhamento de luz (LS), marcados como I_{θ_1} , I_{θ_2} e I_{θ_3} . Através da variação de intensidade do espalhamento são calculados os valores de massa molecular e raio de giro (rg) da molécula detectada (apêndice F). Através das flutuações de intensidade no tempo dos sinais do espalhamento é calculado o raio hidrodinâmico (rh) (apêndice F)

Fonte: Elaborada pelo autor.

As medidas foram feitas com aplicação de 20 μL de amostra com concentração em torno de 4 mg/mL, exceto quando explicitado. Uma quantidade maior que a recomendada pelo fabricante das colunas de exclusão molecular, de 2 mg/mL, foi utilizada porque grande parte das amostras se encontravam sob equilíbrio oligomérico. Portanto, uma concentração maior de proteína poderia aumentar o sinal de populações menos favorecidas de estados oligoméricos.

Foram utilizadas as seguintes colunas de exclusão, de acordo com a disponibilidade das mesmas:

- SEC Protein Column for MALS, 5 μm , 300 $\text{\AA}$, dimensões 4,6 x 300 mm, código WTC-030N5 (Wyatt Technology);
- SEC Protein Column for MALS, 5 μm , 300 $\text{\AA}$, dimensões 7,8 x 300 mm, código WTC-030S5 (Wyatt Technology);

- STR-C SEC-150, 5 μm , 150 Å dimensões 4,6 x 300 mm (Sepax), coluna acoplada a coluna de proteção (coluna guarda) STR-C SEC-150, 5 μm , 150 Å, dimensões 4,6 x 50 mm (Sepax);

A fase móvel e a amostra foram aplicadas utilizando os sistemas Waters 600 Controller e WatersTM 600 Pump (Waters). A saída da cromatografia de exclusão foi acoplada ao detector de espalhamento miniDAWN TREOS (Wyatt Technology) e ao detector de índice de refração Optilab T-rEX (Wyatt Technology). Os ângulos dos detectores de espalhamento são de 43,6°, 90,0° e 136,4° e o comprimento de onda utilizado é de 659 nm. Os dados de variação do índice de refração (dRI) e de espalhamento de luz (LS) foram computados e processados pelo programa ASTRA 7.0.1 (Wyatt Technology). Fundamentos teóricos da técnica estão expostos no apêndice F.

2.7 Cinética por técnica colorimétrica usando *kit* comercial

O experimento foi realizado com o uso do *kit* comercial EnzChek[®] Pyrophosphate Assay (Invitrogen). Esse foi desenvolvido segundo o trabalho de Webb M. R.²³⁵⁻²³⁶ A vantagem desse método foi a praticidade, já que as medidas ocorrem em tempo real não sendo necessário coletar várias alíquotas em diferentes tempos medidos de reação. Além do mais, o protocolo foi realizado em placas de 96 poços, o que automatizou e facilitou as medidas. Algumas das desvantagens da cinética por HPLC são que as alíquotas precisam passar por desnaturação, separação do produto e ainda ser submetida a cromatografia para análise, passos adicionais que acrescentam fontes de erro. Além disso se as proteínas são copurificadas com o produto ele será detectado no HPLC e isso dificulta análises.

Figura 41 - Esquematização ilustrativa dos processos para medição com *kit* comercial EnzChek
Fonte: Adaptada de MOLECULAR...²³⁶

O *kit* contém duas enzimas e um substrato MESG. A primeira enzima é a pirofosfatase inorgânica (IP) que cliva pirofosfato em dois fosfatos inorgânicos. No nosso caso o pirofosfato

é subproduto da formação do c-di-GMP. A enzima PNP combina o fosfato o MESG para formar ribose-1-fosfato e 2-amino-6-mercapto-7-metilpurina, cuja formação pode ser monitorada pela absorção em 360 nm.

A reação utilizada é composta por:

- 5 μ L tampão de reação 20 vezes concentrado (1M tris-HCl, 20mM MgCl₂, pH 7,5);
- 20 μ L de uma solução estoque a 1 mM de 2-amino-6-mercapto-7-metilpurina ribonucleosídeo (MESG);
- 1 μ L de uma solução a 100 U/ μ L de purina nucleosídeo fosforilase (PNP), o que equivale a 1 U;
- cofator MgCl₂, na concentração final de 3mM (1mM do *kit* e 2mM acrescentado de um estoque concentrado 100 vezes);
- De 1 a 3 μ L da enzima que se quer testar a atividade;
- Aditivos como NaCl, c-di-GMP, outra proteína;
- 10 μ L de substrato em água, PPi ou GTP, sendo que a adição desses marcam o início da reação;
- água destilada e deionizada em q.s.p. 100 μ L;

Para um mesmo conjunto de experimentos que se deseja comparar, os reagentes em comuns eram misturados em uma única solução (solução mix) para depois serem distribuídos em uma placa de 96 poços. Uma vez na placa, eram aplicadas as diferentes concentrações de aditivos e/ou enzimas. A placa então fica 10 min temperatura ambiente ($\sim 22^\circ\text{C}$) e logo antes a leitura é adicionado o substrato. As reações têm absorbância acompanhada em leitor de placa Spectramax[®] i3 (Molecular Devices). São realizadas triplicatas para cada condição. Os resultados foram computados e tiveram a primeira etapa de processamento pelo programa SoftMax Pro 6 (Molecular Devices). Nele são armazenados os absorbância e são calculadas as médias e o desvios padrões para as taxas de absorbância pelo tempo (mAU/min), nos intervalos de tempo fornecidos pelo usuário.

As concentrações de substrato e de enzima foram escolhidas através de testes preliminares. Foram testados 0,1 μ M, 0,5 μ M, 1 μ M, 5 μ M e 10 μ M de enzima e concentração GTP fixa de 500 μ M. As concentrações de enzima não podem ficar tão baixas a ponto de fornecer uma reação abaixo do detectável ou com relação sinal/ruído muito baixa. Por outro lado, enzima e substrato, não podem saturar a limitação da velocidade consumo equivalente em pirofosfato apresentada pelo *kit*. Além do mais, a concentração de enzima deve ser desprezível

em relação a concentração do substrato de acordo com o modelo quase-estático de Michaelis-Menten e permitir que o consumo do substrato seja em tempo hábil para serem medidos.

Nos ensaios cinética e de inibição foi estabelecido fixar a concentração de enzima, mutante ou nativa, em 1 a 2 μM . Para estudos de inibição do c-di-GMP foram utilizados a concentração de 100 μM de GTP e concentrações de c-di-GMP fixou-se a concentração de GTP em 100 μM , valor próximo do K_m obtido e valor já encontrado na literatura, e concentração de c-di-GMP de 128 μM a 1 μM . É realizada uma curva padrão como substrato o pirofosfato com concentrações de 64 a 1 μM , preparadas através de diluição seriada. Através da curva padrão foi possível converter absorbância em concentração de GTP, uma vez que para cada GTP consumido é gerado um único PPi ($2\text{GTP} \rightarrow 1\text{c-di-GMP} + 2\text{PPi}$).

Para o cálculo da cinética foram montadas diferentes reações em triplicata com concentrações crescentes de GTP (entre 10 μM e 500 μM). Utilizando as curvas de formação de pirofosfato, foram convertidas as velocidades de formação iniciais (V_0) de absorbância por tempo para consumo de substrato por segundo. Os pontos escolhidos para cálculo de V_0 são regiões 400 segundos até 2000 segundos, na qual a formação foi linear. Tempos iniciais foram evitados pois pode haver contaminante de fosfato diferentes fonte, como do próprio substrato ou do tampão, que irão originar sinal de absorbância. Com os dados de V_0 foi graficado hipérboles retangulares cujos parâmetros do modelo cinético de Michaelis-Menten fornecem a velocidade máxima ($V_{\text{máx.}}$), constante de Michaelis (K_m) e constante catalítica (k_{cat}).²³⁷

Para ensaios de inibição foi utilizada a curva de dose e resposta. No experimento foram calculadas as velocidades de conversão iniciais com a concentração de substrato fixa e variando a concentração de um agente inibitório (I) no meio reacional. No presente estudo, as concentrações de enzima são as mesmas utilizadas na cinética. A concentração de substrato fixada deve ser próxima do K_m . Para DGCs, comumente é usado 100 μM de GTP, concentração utilizada. No presente estudo o próprio produto c-di-GMP foi utilizado como inibidor. O experimento foi feito para estimar o valor de IC_{50} , concentração de inibidor necessária para que a velocidade de conversão do substrato para o produto se reduza em 50 %.

Todos os dados foram processados no programa Origin (versão acadêmica). Os cálculos de propagação de erros foram realizados seguindo normas do material de referência.²³⁸ Os gráficos foram editados no programa Inkscape.²²² Princípios teóricos sobre os cálculos estão expostos no apêndice G.

2.8 Medidas de interação por termoforese em microescala (MST)

O experimento consiste em submeter uma amostra com propriedade fluorescente sobre diferentes concentrações de seu ligante. Caso a migração da proteína fluorescente sobre um gradiente de temperatura seja diferente sobre diferentes concentrações do ligante, essa mudança é atribuída a interação com o ligante (Figura 42).

Para marcação com fluoróforo de um dos parceiros de interação, foram utilizadas duas variantes do NT-647, fluoróforo otimizado pela empresa NanoTemper.²³⁹ Uma das variantes é o NT-647-NHS, ou nhsNT-647, é o NT-647 funcionalizado com o grupo N-hidroxisuccinimida (NHS). O grupo NHS reage com grupos amina, promovendo uma ligação covalente entre NT-647 e a molécula doadora do grupo amina. A segunda variante é o NT-647-maleimida, ou mNT647, no qual o grupo maleimida reage com sulfeto, presente em cisteínas. Nessa reação é formada uma ligação covalente entre o fluoróforo e a molécula regente.

O protocolo de marcação segue as recomendações do fabricante. Para marcação foram destinados 100 μL da proteína que se desejou marcar a 20 μM . Para esses testes as proteínas foram colocadas no tampão hepes (hepes 25 mM, NaCl 300 mM, pH 7,0 ou 8,0), dada a reatividade do grupo NHS com grupos amina, presente no agente tamponante tris. O marcador, em 100 % de DMSO, foi diluído em 100 μL no mesmo tampão da proteína, em quantidade suficiente para concentração 60 μM , 3 vezes a concentração da proteína. A mistura final de 200 μL de proteína e marcador foi homogeneizada por pipetagem e reservada por 30 min no escuro e a temperatura ambiente ($\sim 22^\circ\text{C}$). O próximo passo foi realizar uma dessalinização com coluna de bancada, fornecida pelo fabricante (NanoTemper). Isso para remoção do excesso de marcador. As alíquotas desse procedimento foram quantificadas no 280 nm (calculado pelo ProtParam) e no 650 nm (fluoróforo, $\varepsilon = 250000\text{M}^{-1}\text{cm}^{-1}$).

As medidas foram realizadas no equipamento Monolith NT.115 (NanoTemper). As medidas foram feitas a 25°C , excitação a 20 % a 30 %, e potência média na termoforese (MST). As medidas foram feitas com 20 nM de amostra marcada com adição de 0,05 % de tween 20. Inicialmente são feitos testes iniciais com dois capilares para averiguar a necessidade de utilizar uma concentração de proteína diferente, dada a não saturação da detecção de fluorescência ou a detecção de um sinal abaixo do aceitável pelo equipamento. Nos testes foram verificados também se ocorre aderência da amostra ao capilar, o que indica a necessidade de adicionar na solução BSA 1 mg/mL ou a utilização de capilares tratados específicos.

Em seguida são realizados os procedimentos de *binding check*, dois capilares sem o ligante e dois com a maior concentração possível de ligante, para verificar se ocorria a mudança

na termoforese com a adição do ligante. São realizados então o procedimento de *binding affinity*, o qual capilares contendo diferentes concentrações de ligantes, preparados com diluição seriada, com o intuito de montar uma curva de saturação da mudança de sinal com o ligante. Dessa curva é obtida a curva de constante de dissociação K_d . Fundamentos teóricos da técnica estão expostos no apêndice H e detalhes sobre os cálculos estão no apêndice I.

Figura 42 - Esquematização simplificada da montagem experimental do MST.

Fonte: Adaptada de ENTZIAN; SCHUBERT.²⁴⁰

2.9 Teste de interação com duplo híbrido bacteriano

O presente protocolo foi adaptado¹³ do original descrito no *kit BacterioMatch II Two-Hybrid System* (Agilent). Com o vetor pBT a proteína de interesse é produzida fusionada com o repressor bacteriófago λ (λ cI). O vetor pTRG fornece a proteína presa fusionada com a subunidade α da RNA polimerase (RNAP). Quando a presa interage com a isca, ocorre a estabilização da RNA polimerase na região promotora (*placZ*) e assim ocorre a transcrição dos genes repórteres. O primeiro deles é *HIS3*, que codifica um componente da via biosintética da histidina e o segundo é o *aadA* (ou *Strep^r*), fator que promove resistência da cepa repórter ao antibiótico estreptomicina (Figura 43).

Para co-transformação, foi utilizado 50 μ L de células VR (*BacterioMatch II Validation Reporter Competent Cells*; resistência a canamicina 50 μ g/mL). Para transformação usou-se aproximadamente 80 ng de pTRG e 80 ng de pBT, cada um deles com a região codificante de uma das proteínas que se quer testar a interação. Como controle positivo foi utilizado na co-transformação os vetores pBT-LGF2 +pTRG-Gal11^P. O crescimento seguiu para agitador a

37°C por 16h-24h. A montagem foi feita de manhã para uso das colônias na manhã do dia seguinte.

Figura 43 - Esquematização simplificada do experimento de duplo híbrido bacteriano utilizado.

Se as proteínas de fusão testadas, representadas por A e B, interagem a RNA polimerase é capaz de se estabilizar e transcrever os genes repórteres His3 e Strep^r. Esses desencadeiam a resistência ao inibidor 3-AT e estreptomicina, respectivamente. A resistência ocorre pela O-adenilação no antibiótico, catalisada pela enzima AadA.

Fonte: Elaborada pelo autor.

Com colônias frescas foi preparado pré-inóculos de 4 mL de LB estéril, contendo os antibióticos adequados. Os pré-inóculos dos testes seguiram para agitador shaker por 16h-24h a 30°C. Em uma mesma leva de testes que se deseja comparar, foi tomado o cuidado de misturar os antibióticos ao meio LB para só então distribuir o meio aos tubos de pré-inóculo. Isso para minimizar diferenças de crescimento e diferenças de perdas durante a autoclavagem. Ao final do crescimento, 200 µL do meio de cultivo seguiam foi misturado com glicerol 50 % para estoque no refrigerador a -80 °C. Do restante, 2 mL seguiu para centrifugação a 5000 g por 5 min. Essas células foram resuspendidas em 1 mL de meio mínimo para etapa de lavagem. O resuspenso seguiu para então nova centrifugação e resuspensão, de maneira idêntica à anterior. Esse meio com as células voltou para agitação por 1 h 30 min, com a temperatura de 37 °C. O meio mínimo possui, dentre outros, IPTG, o que faz que as proteínas de fusão serem expressas e, caso interajam, ocorre a expressão também dos genes repórteres His3 e Strep^f.

Ao finalizar esse passo, as células são diluídas em meio mínimo líquido em 1:10, 1:100. As amostras celulares e as diluições são gotejadas em três placas contendo:

- Meio mínimo sólido apenas contendo com 1,5 % (m/vol) de ágar, para controle de crescimento;
- Mesmo meio mínimo sólido com adição de 3-AT (inibidor da via da histidina);

- Mesmo meio mínimo com 3-AT adicionando ainda estreptomicina 12,5 µg/mL. Esse meio é o dual-seletivo. Por verificar a resistência conferida pelo segundo gene repórter *Strep^r* serve ajuda a evitar falsos positivos.

As placas devem seguir para estufa a 37 °C por 24 h e depois ficar a temperatura ambiente e no escuro por mais 24 h, necessário para completar o crescimento de colônias de experimentos com interações fracas. Para registro de imagem das placas finais, essas foram colocadas sobre o orifício de uma câmera escura iluminada por LEDs, elaborada pelo Dr José Fernando Lima.

2.10 Teste de interação por *pull-down*

O experimento foi destinado a testar a interação entre o N-terminal de HsbD (HsbD-Nt) e a porção citoplasmática da PA0847 (HPG-1) *in vitro*, interação já observada no duplo híbrido.¹³ Para esse experimento foi proposto imobilizar a proteína HsbD-Nt fusionada com a 6xHis-SUMO e aplicar a construção HPG-1. Se houver interação, a proteína ficaria retida na resina e seria observada apenas nas eluições. Para o experimento foram preparadas 5 colunas de bancada cada uma contendo 200 µL de resina de cobalto (produto Talon, fabricante Clontech). Todas as colunas foram equilibradas com 20 VC do tampão de purificação padrão (25mM Tris, 300mM de NaCl em pH 8,0). Nas colunas enumeradas de 2 a 5 foi aplicado HsbD-Nt fusionada a 6xHis-SUMO. Aplicou-se a essas colunas lavagens de 20 VC de tampão de purificação padrão, sem imidazol (0mM). Para cada coluna foi realizado então:

- Coluna 1: coluna reservada sem nenhuma proteína imobilizada, mas na qual foi aplicada 10VC do HPG-1 purificada a 0,5 mg/mL;
- Coluna 2: contendo imobilizada a proteína de fusão, que foi reservada para controle;
- Coluna 3: contendo imobilizada a proteína de fusão na qual foi aplicada a aplicada 10VC do HPG-1 purificada a 0,5 mg/mL;
- Coluna 4: contendo imobilizada a proteína de fusão na qual foi aplicada 10 VC do HPG-1 purificada a 0,5 mg/mL e 200 µM de c-di-GMP;
- Coluna 5: contendo imobilizada a proteína de fusão na qual foi aplicada a aplicada 10 VC do HPG-1 purificada a 0,5 mg/mL e 500 µM de GTP e 2 mM de MgCl₂.

Figura 44 - Ilustração experimental do *pull-down*.

No experimento, uma das proteínas que se deseja testar a interação possui afinidade pela resina e a outra não. Caso ocorra a interação, as duas proteínas são eluídas juntas, o que é observado em gel de SDS-PAGE.

Fonte: Elaborada pelo autor.

Em todas as colunas foram aplicados 20 VC do tampão de purificação e feita a eluição com 3VC com o tampão de purificação com 200 mM de imidazol. Aliquotas de cada uma das etapas foram separadas para análise em gel SDS-PAGE.

2.11 Ensaios de cristalização de proteína

2.11.1 Otimização de tampão para ganho de homogeneidade proteica

Antes de se iniciar os protocolos de cristalização, foram realizadas triagem de tampões que melhorassem a estabilidade e a homogeneidade das amostras. O procedimento consistiu em utilizar medidas de espalhamento dinâmico de luz (DLS – *dynamic light scattering*) com diferentes tampões e encontrar algum que aumente a estabilidade proteica. A melhora foi avaliada pelos resultados do experimento para massa (se coerente com a massa esperada) e de polidispersividade da medida (quando menor mais uniforme é a amostra).²⁴¹

Soluções e procedimentos foram baseados no material de referência e é semelhante ao procedimento de otimização de cristalização.²⁴²⁻²⁴³ Resumidamente, foi preparada placa com 24, contendo em cada poço 0,5 mL de cada uma das soluções da triagem, na concentração de 100 mM (Tabela 10). Em seguida, 1 µL da solução de proteína a máxima concentração possível e 1 µL da solução teste são misturados sobre uma lamínula siliconizada. Com a própria lamínula foi lacrado o poço, isso com o uso de silicone. O lado da lamínula contendo a gota ficou voltada para o interior do poço, para alcançasse o equilíbrio de solvente por vaporização. A placa é deixada a temperatura ambiente (22 °C) por 24 h. As gotas que apresentaram ao final desse

tempo a formação de precipitado já foram descartadas. As soluções que não apresentaram precipitado foram utilizadas em medidas no DLS.

As amostras proteicas foram diluídas em 50 mM dos tampões a concentração de $\approx$ 1 mg/mL. Essas amostras ficaram a temperatura ambiente por 10 min. Antes das medidas de DLS, as amostras foram centrifugadas em centrífuga de bancada por 10 min, a temperatura ambiente e na velocidade máxima (16 000 g) para remoção precipitados que podem prejudicar as medidas. O equipamento utilizado foi o Zetasizer Micro V (Malvern Instruments Ltd.). Foi utilizado 20 °C para medidas e a cubeta para 2 μ L de amostra. Os princípios teóricos por trás do DLS seguem o descrito para o espalhamento dinâmico apresentado na seção do SEC-MALS (apêndice F). Foram testadas também medidas com aditivos, como segue:

- 500 mM, 300 mM e 150 mM de NaCl;
- 5 % e 10 % glicerol;
- 1,5 mM DTT;
- 50 mM arginina e 50 mM de ácido glutâmico;
- 5 mM $MgCl_2$;
- C-di-GMP (para construções com domínio GGDEF).

Tabela 10 - Tampões preparados voltados para busca de um que otimize a estabilidade das construções.

Nº	TAMPÃO (100mM)	pH	Nº	TAMPÃO (100mM)	pH
1	Glicina	3,0	13	Cacodilato	6,5
2	Ácido cítrico	3,2	14	Acetato de amônio	7,0
3	-	-	15	MOPS	7,0
4	Ácido cítrico	4,0	16	Fosfato de sódio	7,0
5	Acetato de sódio	4,5	17	HEPES	7,5
6	Fosfato de sódio	5,0	18	Tris	7,5
7	Citrato de sódio	5,5	19	Tris	8,0
8	Fosfato de sódio	6,0	20	Imidazol	8,0
9	Bis-tris	6,0	21	Tris	8,5
10	MES	6,2	22	CHES	9,0
11	Fosfato de sódio	6,5	23	CHES	9,5
12	Bis-tris propano	6,5	24	CAPS	10

Fonte: Elaborada pelo autor.

O tampão e aditivo com melhor ganho na homogeneidade foi utilizado para purificação voltada para experimentos de cristalografia.

2.11.2 Triagem por condições de cristalização

O experimento consiste em testar diferentes soluções de cristalização comerciais para encontrar alguma que promova a formação de cristal proteico, elemento necessário para o prosseguimento em experimentos para obtenção da estrutura. A técnica empregada é a de difusão em gota sentada e as soluções são as dos *kits* INDEX, PEG I-II, MEMB, PEG ION, PEGRX (Hampton Research) e Morpheus (Molecular Dimensions). Foi utilizada para a triagem uma placa de 96 divisões, contendo cada um poço com 70 μL de uma das soluções da triagem e três poços menores onde se adiciona 1 μL da solução e 1 μL de proteína, sendo que para cada um dos poços menores foram testadas concentrações diferentes de proteína. Os poços são lacrados hermeticamente uma vez montada a condição. Algumas das montagens das triagens de condições de cristalização foram feitas manualmente e outras foram realizadas com o sistema automatizado cristalização Crystal Gryphon (ARI), do grupo de Cristalografia (IFSC-USP), com auxílio do pós-doutorando Samuel Guimarães. As placas foram reservadas a 18 °C. Foram testados também, para algumas proteínas, reservar as placas a 4 °C, para diminuir a dinâmica de formação de nucleação de cristais. Cristais podem levar de minutos ou até meses para se formar, portanto foram checadas as condições regularmente.

Uma vez observado cristais, as condições passaram por processos de otimização. Foram montadas caixinhas de 24 poços com 1 mL de soluções com pH, sais e aditivos em torno da condição na qual foi obtida cristais. Gotas de 2 μL da solução do poço são misturadas com 2 μL de solução proteica com diferentes concentrações sobre uma lamínula de vidro siliconizada. O poço é lacrado com ajuda de silicone comercial com a gota voltada para o poço pendurada.

Experimentos de otimização adicionais com aditivos foram realizados com o *kit* Additive Screen™ (Hampton Research) com 96 condições diferentes, *kit* gentilmente cedido pelo Dr. Humberto Pereira. O experimento consiste em utilizar placas de 24 divisões cada uma contendo um único poço com um único suporte para o de difusão de vapor com gota sentada. A abordagem de gota sentada é devido a alteração da tensão superficial que promove o espalhamento da gota para algumas condições. No poço é adicionado na solução que promove a cristalização enquanto que na concavidade são adicionados 1 μL do aditivo, 4 μL da solução do poço e 5 μL da solução proteica. Os poços são selados e mantidos a mesma temperatura da condição de cristalização.

2.11.3 Expressão com selenometionina

Uma das opções para resolução de estruturas cristalográficas proteicas é usar o sinal anômalo de difração de raios-X provindo do átomo de selênio da proteína de interesse cristalizada. Para isso, foi realizada a substituição das metioninas naturais por selenometioninas, através da expressão em meio mínimo. A difração resultante dos experimentos então é utilizada para resolução através do técnicas que se aproveitam do sinal anômalo utilizando um único comprimento de onda (SAD) ou de vários (MAD) (abreviações respectivamente para *single-wavelength anomalous diffraction* e *multi-wavelength anomalous diffraction*).

Para obter a proteína com selenometionina incorporada, procedeu-se com a expressão heteróloga padrão com o meio mínimo M9 enriquecido com aminoácidos naturais, excluindo a metionina, e selenometionina. No preparo do meio M9 devemos diluir em 970 mL de água destilada, para cada litro que se pretende expressar:

- 6,8 g de fosfato de sódio;
- 3,0 g de fosfato de potássio monobásico;
- 0,5 g de cloreto de sódio;
- 2,0 g de cloreto de amônio.

A solução deve ser autoclavada e apenas logo antes ao uso deve ser acrescido as soluções devidamente estéreis através de filtração com filtros 0,22 µm:

- 2 mL de sulfato de magnésio a 1 M;
- 2 mL de cloreto de cálcio a 0,1 M;
- 25 mL de glicose 40 % (m/vol);
- 1 mL de tiamina 10 mg/mL.

Ao final se obtém 1 litro de meio M9. É necessário separar 100 mL desse meio em um frasco, também esterilizado previamente, para uso como pré-inóculo. O primeiro passo para o preparo do pré-inóculo é recolher 5 mL do frasco contendo 100 mL de M9 e transferi-lo uma placa com células de expressão, com o auxílio de uma alça de Drigasky. O resuspenso é transferido de volta para o pré-inóculo. Com antecedência, portanto, deve-se fazer a transformação da célula de expressão, aqui utilizado a cepa BL21(DE3) com o plasmídeo contendo o gene de interesse em meio LB ágar e os antibióticos de seleção, que no caso do vetor pETSUMO é requerido canamicina a 50 mg/mL.

O pré-inóculo deve seguir para agitação a 37 °C e 250rpm por volta de 1h a 1 h 30 min. Antes da transferência do pré-inóculo para o inóculo foi necessário separar 10 mL do meio mínimo para dissolver os aminoácidos e por volta de 2 mL para medidas de D.O. Inocular o restante da solução com o pré-inóculo e continuar o cultivo a 37 °C e 250 rpm para a multiplicação celular. Enquanto o meio não atinge a D.O. entre 0,4 e 0,6 são dissolvidos os aminoácidos no meio M9 reservado:

- 100 mg de L-lisina;
- 100 mg de L-fenilalanina;
- 100 mg de L-isoleucina;
- 50 mg de L-leucina;
- 50 mg de L-valina;
- 60 mg de selenometionina.

Atingida a D.O. adequada o meio continua sob as mesmas condições de agitação e temperatura por mais 15 minutos. E só então é acrescido 0,1mM de IPTG e a temperatura do agitador é diminuída para 18 °C. A partir dessa etapa se prossegue com o protocolo de purificação usual, como descrito na seção de purificação 2.3.2.

2.11.4 Experimentos de difração de raios X

Os cristais foram submetidos a difração de raios X na linha MX-2 no Laboratório Nacional de Luz Síncrotron (LNLS, Campinas – SP). Uma coleta foi realizada no síncrotron do *Diamond Light Source*, em Oxford, Inglaterra, pelos doutores Bruno Matsuyama e Flávio Rosetto. Antes de serem submetidos ao feixe de raios X, os cristais foram submersos em soluções crioprotetores, que tiveram a finalidade de evitar a formação de cristais de gelo no congelamento para difração e para melhora na qualidade do padrão de difração. Essas soluções eram polietilenoglicol (PEG) 400 e 200, etilenoglicol e glicerol, todos nas concentrações testadas entre 10 % e 25 %. Os dados das coletas foram processados nos programas iMosflm²⁴⁴ e XDS.²⁴⁵

2.12 Análise em larga escala de sequências de proteínas com GGDEF e EAL

Para obter indícios de como é a dinâmica entre GGDEFs e EALs de proteínas que possuem esses dois domínios em série, foram realizadas análises de diversas sequências assim classificadas. Para isso, foram utilizados dois bancos de dados. O primeiro banco é referente ao grupo de Galperin e colaboradores.^{91,200} e o outro é a respeito do banco de dados do Uniprot.

O banco de Galperin e colaboradores^{91,200} é uma compilação de sequências de proteínas das vias do c-di-GMP, selecionadas do banco Uniprot e categorizadas. São sequências detentoras, de pelo menos um, dos domínios GGDEF, EAL, HD-GYP, PilZ e *riboswitches* I e II. Na página *online* essas sequências estão apresentadas de acordo com os organismos de origem e separados de acordo com a presença do GGDEF, do EAL ou se ambos estão na mesma cadeia polipeptídica. O segundo banco utilizado é a totalidade do Uniprot. Esse grupo abrange diversas sequências com diferentes graus de confiabilidade, medida através do número de revisões. Esse banco foi usado pelo elevado número de sequências apresentados, requisito para obtenção de uma relevância estatística maior.

Tanto na obtenção, quanto na grande parte do processamento das sequências, foram aplicados códigos em python. Além dos comandos comuns de armazenamento e processamento de *strings*, foram aplicadas funções da biblioteca *urllib* para coleta de dados na rede e de expressões regulares (*regex*) tanto para facilitar a extração de dados *online* quanto para o processamento das sequências. Do banco disponibilizado pelo grupo de Galperin foram excluídas as sequências com links para o Uniprot obsoletos e com problemas na leitura do html. Do grupo de sequências do Uniprot foi necessário apenas utilizar os códigos do Pfam na página principal para recuperar uma lista de sequências com os domínios desejados.

Das sequências recuperadas com êxito, foram alinhadas e extraídas as regiões identificadas como GGDEF e EAL. Para isso foram usados comandos do programa *hmmer-3.1b2*. Para isso foi necessário primeiro computar os modelos de *hmm*, através do comando *hmmbuild*. Nesse procedimento foi fornecido o *seed* (alinhamento) do Pfam que define o domínio. Isso foi feito, separadamente, para cada um dos domínios em estudo. Com o perfil de *hmm* foi possível utilizar o comando *hmmalign* com as opções *--trim* e *-o* no conjunto de sequências em estudo. As opções desse comando são para a seleção apenas da região a qual foi identificado o domínio nas sequências e a armazenagem das mesmas em um arquivo *.sto* (formato *Stockholm*). Linhas de código foram elaboradas para recuperar parte das sequências

conectoras entre GGDEF e EAL eliminadas nos comandos do HMMER. As regiões conectoras então foram alinhadas com o programa ClustalW, utilizando opções padrão. Outras funções em Python foram escritas para juntar os alinhamentos do domínio GGDEF, da região conectora e do EAL em um único alinhamento múltiplo (ou MSA, abreviação para *multiple sequence alignment*).

Além disso, linhas de código foram utilizados para filtrações extras, para garantir a qualidade das sequências. Os filtros excluíram, especificamente sequências nas quais:

- a distância entre GGDEF e EAL identificados pelo HMMER é maior que 180 resíduos;
- inseriam espaçamentos no alinhamento (*gaps*) na série de motivos conservados (motivo) do sítio ativo (sítio A);
- possuíam um número de espaçamentos (*gaps*) no alinhamento, tanto do EAL quanto do GGDEF, maior que 200;
- sequências que inserem espaçamentos no motivo GGDEF do domínio GGDEF;
- nas quais o GGDEF vem depois do EAL;
- nas quais a sequência correspondente ao GGDEF e EAL se sobrepunham.

Através de códigos em Python foi possível automatizar a identificação e agrupamento das sequências de acordo com a conservação dos domínios GGDEF e EAL e também do motivo RXXD. Essa classificação foi feita através verificação da presença de alguns dos resíduos característicos desses elementos, dada a correspondência das posições desses resíduos no alinhamento final, utilizando as sequências referência de PleD e TBD1265.

Para classificar o domínio GGDEF como conservado, foram checados a presença do aspartato que coordena Mg^{2+} e outro que confere especificidade pelo grupo guanil. Esses resíduos correspondem a D327 e D344 da proteína de referência PleD. As posições correspondentes ao motivo GGDEF, grupo de cinco resíduos que confere nome ao domínio, também foram consideradas. Para primeira posição, G368 na referência PleD, foram aceitos, além da glicina (G), alanina (A) ou serina (S). Isso devido as exceções que vem surgindo na literatura de DGCs ativas a respeito dessa posição. Na segunda posição do motivo GGDEF, foi aceito apenas uma glicina. Para terceira posição, o resíduo essencial da catálise, foram aceitos tanto um aspartato quanto um glutamato (resíduos D e E). O quarto resíduo do motivo GGDEF, utilizado na coordenação do cátion divalente, foi verificada a presença de um glutamato. O último resíduo do motivo foi requerido para conservação, ou uma fenilalanina ou leucina.

Para o sítio inibitório foi considerado a presença do motivo característico RXXD, que corresponde a presença dos resíduos R359 e D362 de PleD. Uma das hipóteses é que proteínas

contendo tanto o domínio GGDEF e EAL, o mecanismo de regulação da atividade GGDEF não ocorreria através do sítio inibitório e por isso esse sítio estaria ausente nessas proteínas

Para classificação de uma sequência com EAL conservado foram analisadas a presença dos resíduos da coordenação de um dos cátions divalentes, o primeiro glutamato (E) da assinatura EAL, que corresponde a E523 da referência TBD1265, além dos resíduos N584, E616 e D646. Foi considerado ainda o resíduo que coordena o segundo cátion divalente D647. Apesar de não ser um dos resíduos essenciais identificados em RocR, D647, juntamente com D646, compõe os primeiros resíduos do motivo regulatório D⁶⁴⁶DFG(T/A)GYSS, foi considerado para classificação do EAL conservado. Como visto (seção 1.9.2), esse motivo aparece essencial na dimerização e regulação da atividade de EALs e o resíduo. E ainda, outros resíduos considerados para classificar uma sequência com domínio EAL conservado foram E619 e E703, resíduos descritos estabilizador do loop 5 e indicado como catalítico, respectivamente.

Tabela 11 - Posições e conjunto as opções de resíduos que devem ocupá-las para classificar como conservado os domínios ou o motivo.

Domínio GGDEF:	Resíduos D-D	327: D; 344: D
	Motivo GGDEF	368: ou G ou A ou S; 369 :G; 370 :D ;371: E, F372: F ou L
	Motivo RXXD	359: R; 362: D
Domínio EAL:	E-E-N-D	523: E; 584: N; E616:E, 646: D
	DDFG(T/A)GYSS	647: D
	E catalítico	703: E
	E loop	619: E

* PleD é a referência para o domínio GGDEF (que engloba o motivo RXXD); TBD1265 é a referência para o domínio EAL.

Fonte: Elaborada pelo autor.

O alinhamento múltiplo resultante do processamento de sequências dos bancos de Galperin e do Uniprot também foram submetidas a análises de coevolução de resíduos, através de programas compilados no programa PFstats (abreviação para *protein families using simple statistics*), elaborado por Bleicher e colaboradores. Esse tipo de abordagem é muito útil na predição de contatos devido ao enovelamento, sítios funcionais, contatos entre proteínas distintas e conformações alternativas.²⁴⁶ Detalhes sobre os cálculos realizados com o PFstats estão expostos no apêndice J.

Para complementar as análises feitas pelo PFstats, linhas de código foram montadas para automatizar o processamento das frequências de cada um dos resíduos de cada posição do alinhamento para tentar relacionar as frequências dos resíduos com a conservação dos domínios GGDEF e EAL. Outros códigos foram desenvolvidos para automatizar a desenho das redes de correlação e da conversão das posições do alinhamento em posições das sequências de referências tomadas, que são o GGDEF de PleD e o EAL de TBD1265.⁸²

3 RESULTADOS E DISCUSSÕES

3.1 Estudos funcionais das diguanilato ciclases PA0847 e HsbD

Nessa seção serão descritas e discutidas as caracterizações experimentais *in vitro* das construções de PA0847 e da região N-terminal de HsbD. Partindo destes resultados, serão apresentados potenciais modelos de mecanismos de regulação desempenhada pelos domínios de PA0847 e sobre a participação de HsbD nesses mecanismos.

3.1.1 Planejamento experimental e produção heteróloga das proteínas do trabalho

A estratégia adotada para o estudo das proteínas sensoras PA0847 e HsbD (PA3343) e outras (apêndice A) focou na caracterização de fragmentos periplasmáticos ou citoplasmáticos. Os bons progressos em experimentos com PA0847, associados as descobertas inéditas da interação forte entre PA0847 com HsbD¹³, motivaram focar trabalhos com essas proteínas *in vitro*. Como descrito na seção 2.1, o uso de construções foi adotado, dentre outras, dada as dificuldades técnicas de solubilização de hélices hidrofóbicas que, *in vivo*, estão embebidas na membrana. Várias construções foram desenhadas para maximizar as chances de obter alguma solúvel. Com os dados de predição de domínios, estruturas secundárias e de regiões transmembrana, foi possível planejar as regiões das proteínas que seriam trabalhadas *in vitro* (Figura 45). Os limítrofes das construções foram determinados, sempre que possível, em regiões sem predição de elementos de estruturas secundárias regulares, hélices alfa e fitas beta, pois isso poderia promover instabilidade na proteína truncada final.

Figura 45 - Exemplificação do planejamento de múltiplas construções aplicados para o domínio CHASE4 de PA0847. O diagrama central mostra a predição de estruturas secundárias, fornecido pelo PSIPRED. Nesse, as hélices estão em rosa e as fitas em amarelo. As setas acima do diagrama das estruturas secundárias fazem referência às abrangências de regiões preditas transmembrana (TM) e domínios proteicos, de acordo com o servidor TMHMM e Pfam, respectivamente. No exemplo, determinou-se que o melhor ponto para início de todas as construções seria o resíduo D65, como indicado nas setas abaixo do diagrama de predição de estruturas secundárias. Já para a posição final, foram escolhidos pontos que delimitariam construções contendo a região predita do domínio CHASE4 (resíduos M240 e M247, construções C4-1 e C4-2 respectivamente) ou toda a região periplasmática (por exemplo, resíduo D301, construção C4-3).

Fonte: Elaborada pelo autor.

A Figura 48 resume os dados de predição dos domínios, hélices transmembrana e as respectivas construções propostas para PA0847 e HsbD. Os detalhes de cada uma das construções se encontram na Tabela 12).

As sequências dos domínios foram identificadas, separadas e alinhadas com sequências dos mesmos domínios de outras proteínas já abordadas na literatura. Essa análise é essencial para verificar a presença de resíduos já descritos como importantes para a função. Uma dessas análises motivaram, por exemplo, considerações sobre questões funcionais em relação ao domínio HAMP no planejamento das construções de PA0847 (Figura 46). Estudos com proteínas sensoras das famílias HK e MCP demonstraram que modificações no domínio HAMP fornecem proteínas com atividades distintas. E isso foi correlacionado com mudanças conformacionais do domínio, como descrito na seção 1.13.4.^{130,183,186,188}

Utilizando o sensor de serina Tsr como referência, seriam ativas as construções HPG-1, que engloba todo domínio HAMP, e HPG-3, que possui apenas a hélice AS2 do domínio. Além disso, também forneceria atividade as construções PG-1 e PG-2, que possui apenas o domínio GGDEF e PAS. Enquanto isso, HPG-2, que não possui a hélice AS1, seria inativa (Figura 46).

Figura 46 - Construções de PA0847 e o padrão do impacto da remoção das porções do domínio HAMP em proteínas inteiras de Tsr, sensor quimiotático de serina.

O alinhamento mostra, além da sequência HAMP da Tsr, a sequência de HAMPs identificada em PA0847. Em vermelho, estão em destaque posições conservadas de HAMPs. Cada linha abaixo do alinhamento resume experimentos *in vivo* com o sensor inteiro de serina Tsr. Modificações foram feitas no HAMP em duas cepas, uma superexpressando CheB e CheR (*cheRB*⁺) e outra com essas proteínas deletadas (*ΔcheRB*). Essas proteínas são enzimas autoregulatórias de sistemas de sinalização MCP (Figura 22, seção 1.12). As barras horizontais indicam as abrangências do domínio HAMP que foram removidas de Tsr. Os gráficos de barras à direita mostram os resultados de ativação do flagelo, atribuída a ativação do sistema de sinalização. As setas coloridas indicam a parte inicial de construções de PA0847: HPG-1 (roxo), HPG-2 (azul), HPG-3 (verde azulado) e PG-1 e PG-2 (verde).

Fonte: Adaptada de AMES; PARKINSON.¹⁸⁸

A partir do alinhamento dos domínios GGDEF referência, constata-se que PA0847 e HsbD possuem o sítio I, o sítio A e todos os resíduos característicos para catálise (Figura 47).^r De fato, era esperado que essas proteínas fossem ativas, devido a resultados de produção de c-di-GMP em resposta a superexpressão, como visto por Kulesekara e colaboradores¹⁹³ e outros.⁸ Por isso, foi produzido um mutante R587A da construção HPG-1 de PA0847, referente ao sítio inibitório RXXD. A construção com a mutação foi denominada HPG-1^{AXXD}. O resíduo mutado corresponde a R359 de PleD, conhecido por participar das duas conformações ligadas ao c-di-GMP (seção 1.8.2, Figura 13 e Figura 14). Com o mutante, seria possível observar o impacto da ausência da ligação do c-di-GMP para atividade e para alternância de estados oligoméricos empregados na ligação ao dinucleotídeo.

Decidida as construções e a abrangência das mesmas, seguiu-se o planejamento de oligonucleotídeos iniciadores (*primers* - apêndice B). Com base nos critérios descritos acima, foram delineadas um total de 14 construções, sendo dessas 13 de PA0847 e 1 de HsbD, para

^r Para consultar o alinhamento completo, consulte o apêndice C, Figura 90.

serem produzidas de modo heterólogo, visando caracterizações estruturais e biofísicas (Figura 48).

Figura 47 - Parte do alinhamento entre GGDEFs das proteínas em estudo com outras de referência.

É mostrado parte das sequências GGDEF alinhadas de PA0847, HsbD e PA1851 (similar a HsbD) com os GGDEFs de referência PleD, DgcZ, WspR e YfiN. A numeração é referente a PleD. Em vermelho são destacados alguns dos resíduos característicos do domínio GGDEF, como descrito na seção 1.8. São importantes estruturalmente a volta conservada DXLT e o resíduo R366, esse último usado para estabilização de um grampo formado pelo sítio A. Alguns dos resíduos que são empregados na catálise são D327, N335, D344 e os resíduos da assinatura GGD(E)EF. Outros resíduos, como aqueles que compõem o sítio inibitório primário Ip e o secundário 1, Is (1), participam das conformações inibidas do GGDEF (seção 1.8.2).

Fonte: Elaborada pelo autor.

Figura 48 - Resumo detalhado das arquiteturas de domínios e das construções 13 construções de PA0847 (A) e da construção da porção N-terminal de HsbD (B). As construções estão representadas por setas coloridas, abaixo dos esquematizações de cada arquitetura de domínio das proteínas. As construções de PA0847 foram denominadas de acordo com suas composições de domínios (CHASE – C4; HAMP – H; PAS – P; GGDEF – G). Nos casos em que diferentes construções englobassem o(s) mesmo(s) domínio(s), as construções receberam um número. A numeração acima das arquiteturas das proteínas representam a posição nas sequências dos limites dos domínios preditos no servidor PFAM (preto) ou hélices transmembranares preditas no servidor TMHMM (vermelho). A numeração encontrada abaixo das linhas pontilhadas que delimitam as construções indicam as respectivas posições nas sequências. A princípio, levando-se em conta apenas a arquitetura de domínios de PA0847, apenas 6 construções (C4, HPG, HP, PG, P e G) seriam suficientes para cobrir todas as possibilidades de domínios para PA0847, excluindo uma construção do domínio HAMP isolado, altamente instável em solução. Entretanto, como visto, múltiplas construções, envolvendo os mesmos domínios, visaram o aumento da probabilidade da obtenção de amostras solúveis.

Fonte: Elaborada pelo autor.

Usando as sequências de aminoácidos das construções foram calculados parâmetros com o ProtParam, alguns deles listados na Tabela 12. Esses dados foram úteis nos processos de obtenção e manipulação das amostras.

Tabela 12 - Detalhes das construções planejadas para trabalhos em solução e alguns de seus respectivos parâmetros.

	Construções	Domínios	Pares de base	Resíduos e limites ^(*)	Massa (kDa) ^(**)	pI ^(**)	ϵ mg ⁻¹ .mL ^(**)
PA0847	C4-1	CHASE4	528	176 (65-240)	20,2	4,5	1,04
	C4-2	CHASE4	549	183 (65-247)	21,0	4,7	1,00
	C4-3	CHASE4	711	237 (65-301)	27,0	5,1	1,03
	C4-4	CHASE4	747	249 (65-313)	28,5	5,4	1,03
	C4-5	CHASE4	744	248 (65-312)	28,4	5,5	1,04
	HPG-1	HAMP+PAS+GGDEF	1020	340 (339-678)	38,2	5,3	0,65
	HPG-1 ^{AXXD}	HAMP+PAS+GGDEF	1020	340 (339-678)	38,2	5,3	0,65
	HPG-2	HAMP+PAS+GGDEF	963	321 (358-678)	35,9	5,2	0,54
	HPG-3	HAMP+PAS+GGDEF	930	310 (369-678)	34,6	5,1	0,56
	HP	HAMP+PAS	492	164 (339-502)	18,6	5,2	0,70
	PAS	PAS	303	101 (402-502)	11,2	5,0	0,66
	PG-1	PAS+GGDEF	831	277 (402-678)	30,9	5,1	0,63
	PG-2	PAS+GGDEF	825	275 (404-678)	30,5	5,7	0,63
	G	GGDEF	303	162 (517-678)	18,3	5,7	0,57
HsbD (PA3343)	Nt	Região N-terminal	153	51 (1-51)	6,2	7,9	0,70

* Os primers usados para as clonagens estão listados nas tabelas do apêndice B, Tabela 26.

** Dados computados pelo programa *online* Proparam²⁰⁸, utilizando a sequência das respectivas construções.

Fonte: Elaborada pelo autor.

3.1.2 Clonagem, expressão e purificação das construções de PA0847

Após otimizações em diferentes etapas no processo de clonagem das construções no vetor de expressão pETSUMO^s, todas as construções planejadas foram concluídas e confirmadas por sequenciamento (seção 2.2). (Figura 49).

Para a obtenção do vetor de expressão da construção HPG-1 contendo a mutação R587A (HPG-1^{AXXD}), foi utilizado o protocolo de PCR de mutagenese sítio-dirigida (seção 2.2.7). Esses experimentos foram conduzidos com um par de oligonucleotídeos adequados (apêndice B, Tabela 28) e o plasmídeo pETSUMO-HPG-1 como molde. O vetor pETSUMO-HPG-1^{AXXD} foi obtido após uma extensa otimização das reações, que incluíram a adição de DMSO, o uso de diferentes temperaturas de hibridização e o uso das polimerases Pfu e Phusion (Thermo Scientific). (Figura 50).

^s Em vários casos, a adição de BSA 1,5 mg/ml elevou o rendimento da amplificação da região de interesse na PCR. No geral, as clonagens que usaram as endonucleases BamHI e HindIII necessitaram passar pelos procedimentos uma única vez para serem finalizadas (Tabela 26, apêndice B).

Figura 49 - Análise de restrição em gel de agarose das construções de PA0847 e PA3343 em pETSUMO.

Como explicitado na seção de materiais e métodos (seção 2.2), as regiões codificantes das construções foram alocadas no vetor pETSUMO, que possui aproximadamente 5,7 kpb. As marcações acima dos géis representam o padrão de massa molecular utilizado (MM) e a denominação da construção analisada, de acordo com o padrão de nomenclatura apresentado na Tabela 12 e na Figura 48. A altura da banda correspondente ao vetor clivado está em destaque, ao lado das fotos dos géis. As bandas correspondentes a cada um dos CDS estão indicadas com uma seta vermelha. Os tamanhos das CDSs estão resumidos na Tabela 12.

Fonte: Elaborada pelo autor.

Figura 50 - Análise em gel de agarose de amostras das PCRs para obtenção de pETSUMO-HPG-1^{AXXD}.

Nesse experimento é realizada a amplificação do vetor inteiro. A análise em gel é uma etapa opcional no procedimento, pois o veredito do sucesso da PCR vem com a presença de colônias transformantes da reação. Foram testadas as enzimas Pfu e Phusion. Em todas as condições testadas com Pfu observar-se a presença da banda do amplificado, o que revela a maior eficiência da mesma. A altura do plasmídeo amplificado contendo a CDS, assim como a altura correspondente ao excesso de primers, estão destacadas ao lado do gel, em vermelho.

Fonte: Elaborada pelo autor.

Todas as construções de PA0847 foram submetidas aos mesmos testes de expressão e purificação, como descrito nas seções 2.3.1 e 2.3.2. A Figura 51 apresenta de forma resumida, os resultados das etapas de purificação, usando como exemplo as construções C4-1 (Figura 51A) e HPG-3 (Figura 51B). Prosseguiu-se os trabalhos com C4-1, C4-2, HPG-1, HPG-2, HPG-3, HP, P, PG-1 e G. Essas construções foram obtidas solúveis e com alto grau de pureza. Amostras finais das construções estão apresentadas na Figura 52.

Figura 51 - Resultados das etapas de purificação aplicadas para construções da PA0847.

A) Exemplificação com a construção C4-1. A proteína em fusão com a SUMO possui aproximadamente 33,8 kDa, enquanto é esperado para a construção C4-1 uma massa de 20,2 kDa (Tabela 12); B) Mesmos procedimento realizados para proteína HPG-2. Em fusão, a construção possui 48,2 kDa, enquanto que isolada, é esperado a massa de 34,6 kDa. Abreviações: P – *pellet* (fração insolúvel); S – sobrenadante; NL – fração não ligada (da cromatografia de afinidade); CV – resultado do clivado com Ulp (SUMO protease); 0mM , 5 mM, 10 mM e 500 mM – lavagens e eluição com o tampão de purificação com as respectivas concentrações de imidazol na primeira afinidade; L1 e L2 – lavagens realizadas na segunda etapa de afinidade. Apesar da massa da fusão SUMO ser 13,6 kDa, ela migra no gel SDS com tamanho próximo ao do marcador de 20 kDa.

Fonte: Elaborada pelo autor.

Figura 52 - Resultado do SDS-PAGE 15 % das amostras finais das construções solúveis de PA0847.

Fonte: Elaborada pelo autor.

As construções C4-3 e C4-4, que englobam quase a totalidade da porção periplasmática de PA0847, se revelaram em grande parte na fração insolúvel, após o rompimento celular. Essas construções periplasmáticas apresentaram baixos rendimentos de proteína instável, menores

que 1 mg por litro de cultura. Devido aos bons resultados com C4-1 e C4-2, otimizações adicionais nos protocolos de C4-3, C4-4 não foram executadas. Devido à alta semelhança entre as construções C4-4 e C4-5, esta última não foi priorizada para testes.

Devido a baixa eficiência da clivagem da fusão 6xHis-SUMO da construção PG-2 e tendo em vista a alta semelhança entre PG-1 e PG-2^t, a construção PG-2 foi descartada. Portanto o trabalho foi conduzido somente com a construção PG-1, que será abreviada a partir de então apenas como PG.

Tabela 13 - Resumo dos resultados nas etapas de produção e rendimento médio das preparações para PA0847.

Construções	PCR	Clonagem	Expressão	Solubilidade	Purificação	Rendimento*
C4-1						30±2
C4-2						44±1
C4-3						-
C4-4						-
C4-5						-
HPG-1						4,4±0,6
HPG-1 ^{AXXD}						1,6±0,2**
HPG-2						11±5
HPG-3						20±6
HP						13±4
PAS						10±6
PG-1 (PG)						25±2
PG-2						-
G						18±5

Em azul estão as etapas bem-sucedidas, em amarelo estão etapas cumpridas com alguma dificuldade e em vermelho estão as etapas que precisariam ser otimizadas.

* Médias baseadas em três ou mais purificações independentes (medidas logo após a última cromatografia de afinidade);

** Apenas na expressão heteróloga da HPG-1^{AXXD} em BL21(DE3), por vezes, o *pellet* celular final apresentou-se mais escurecido em relação ao padrão. Esse fornecia rendimento praticamente nulo, quando submetido a etapas de purificação. A possibilidade do problema ser causado por uma mutação ou deleção no vetor foi descartada, pois a região codificante de PA0847-HPG-1^{AXXD} foi extraída e reinserida em outra amostra de pETSUMO. Algumas hipóteses foram levantadas, tais como a sensibilidade a produção de c-di-GMP pela construção, a temperatura de indução, ao uso de colônias recém transformadas ou na saturação de células. Porém tais fatores não foram investigados.

Fonte: Elaborada pelo autor.

A obtenção de algumas amostras instáveis era algo esperado, justificando a estratégia adotada com várias construções, mesmo com tamanhos muito próximos. Grande parte das construções foram obtidas e com alta solubilidade o que mostra a eficácia do planejamento das construções. A Tabela 13 resume o progresso dos experimentos das construções de PA0847 assim como os respectivos rendimentos finais.

^t PG-2 possui apenas dois resíduos a menos no N-terminal em relação a PG-1 (Figura 48 e Tabela 12).

Foi observado que o rendimento em massa de proteína por litro de cultura foi inversamente proporcional a eficiência da mesma em produzir c-di-GMP (como será visto na seção 3.1.5). Provavelmente, o c-di-GMP sintetizado pelas construções durante a expressão heteróloga poderia estar interferindo no metabolismo da célula de expressão. Esse fato não foi investigado a fundo.

Desde os primeiros testes de purificação, foram observados múltiplas populações eluídas na cromatografia de exclusão das construções citoplasmáticas puras no geral, exceto para construção P. Amostras dos picos de eluição, quando concentrados e recromatografados, forneciam picos de eluição nos mesmos volumes, reforçando que se trata da mesma amostra em diferentes populações oligoméricas e/ou conformacionais (Figura 53A). Esse comportamento será extensamente explorado nas próximas seções dessa tese (subseções da seção 3.1.4).

Figura 53 - Cromatogramas de amostras da HPG-1, ilustrando o equilíbrio oligomérico (A) e o ganho de resolução com diferentes colunas de exclusão molecular (B).

A) Na recromatografia, utilizando a coluna Superdex 75 10/300 (GE) de qualquer um dos picos de HPG-1, marcados com *, é gerado um perfil idêntico ao inteiro original; B) Comparação entre cromatogramas de HPG-1 em diferentes colunas de exclusão molecular exemplificando o ganho de resolução das diferentes populações com a Superdex 200 10/300 GL (GE). Por isso essa coluna foi utilizada para os subsequentes experimentos.

Fonte: Elaborada pelo autor.

Uma melhor resolução das populações foi obtida usando a coluna de exclusão molecular Superdex 200 10/300 GL (GE). Por isso, essa coluna foi escolhida como padrão para realização das purificações e análises posteriores. No exemplo com HPG-1 (Figura 53B), na coluna

Superdex 75 10/300 GL (GE) foi possível separar duas populações, enquanto que na Superdex 200 10/300 GL (GE) foi possível resolver 3 populações.

3.1.3 Clonagem, expressão e purificação da construção N-terminal de HsbD

Assim como para as construções de PA0847, os procedimentos de clonagem foram aplicados para região N-terminal da proteína HsbD (HsbD-Nt) no vetor pETSUMO, a partir do DNA genômico de *P. aeruginosa* (seção 2.2), (Figura 54).^u Resultados de sequenciamento confirmaram a sequência correta de HsbD-Nt.

Figura 54 - Gel de análise de restrição de HsbD em pETSUMO.

Como explicitado na seção de materiais e métodos, as regiões codificantes das construções foram alocadas no vetor pETSUMO possui aproximadamente 5,7 kpb. O tamanho da sequência de interesse é de 151 pb, não incluindo a região adicional dos primers.

Fonte: Elaborada pelo autor.

Desde os primeiros testes de expressão, análises de SDS-PAGE demonstraram um padrão recorrente de duas bandas da proteína expressa, com massas moleculares compatíveis com as massas esperadas para 6xHis-SUMO-HsbD-Nt. Embora sempre presente, a intensidade das bandas nesse padrão alterava em resposta a mudanças das condições de expressão, tais como cepa de *E. coli* utilizada para expressão e a temperatura, indicando que ambas eram produto do vetor de expressão, e não de um contaminante endógeno (Figura 55A). Esse padrão de duas bandas também foi observado para a região N-terminal da proteína PA1851, uma proteína semelhante a HsbD e que interage de forma fraca com PA0847.¹³

Levantou-se a hipótese para as duas bandas expressas de HsbD foi consequência de degradação por proteases da *E. coli* utilizada para expressão. Entretanto, testes com a adição com o inibidor de proteases PMSF (fluoreto de fenil-metano-sulfonil) ou com coquetel de

^u A adição de BSA 1,5 mg/ml na PCR foi crucial para observação da amplificação da região de interesse. Reações de ligação com proporção maior de inserto, de 6:1 e 12:1 (proporção CDS:vetor) na reação de ligação, resultaram em maior eficiência na transformação.

inibidores de proteases complete EDTA-free (Roche), no tampão de purificação não alteraram os resultados. Uma outra hipótese foi a de interrupção prematura da transcrição, devido à presença de códonos menos utilizados por *E. coli* da CDS de HsbD-Nt. De fato, um enriquecimento da banda de maior massa molecular do padrão de expressão foi observado em testes com a cepa de *E. coli* Rosetta, principalmente quando a expressão era realizada à 18 °C. Nesse sentido, testes foram realizados utilizando CDS sintética de HsbD-Nt otimizado para expressão em *E. coli* (adquiridos pela empresa GenScript no vetor pUC57, apêndice B - Tabela 27). Essas CDS que foram transferidos para pETSUMO. Porém, os resultados obtidos com esse clone foram similares (Figura 55A).

Figura 55 - Análises em SDS-PAGE e cromatograma final das etapas de purificação de HsbD-Nt.

A) Géis dos testes de expressão em BL21(DE3) e Rosetta; B) Resultado de clivagem com Ulp, com as duas bandas obtidas destacadas com setas e análise do precipitado, mostrando a presença de HsbD-Nt. Cromatograma da exclusão molecular depois da clivagem da fusão 6xHis-SUMO-HsbD-Nt, acompanhada do resultado do gel SDS-PAGE dos picos de eluição coletados.

Fonte: Elaborada pelo autor.

Durante a etapa de clivagem com Ulp, grande parte da proteína precipitou. Esse precipitado, quando analisado em gel, revela uma banda compatível com a massa esperada para HsbD-Nt (figura 52B). E isso ocorre mudando o pH do tampão purificação de 8,0 para 7,0,

dado que HsbD possui pI predito de 7,9 (Tabela 12). Após clivagem da proteína de fusão, o padrão de duas bandas em solução se mantém (Figura 55B), com massas moleculares próximas ao esperado para HsbD-Nt, de aproximadamente 6,5 kDa (Tabela 12).

Para contornar o problema de precipitação, foi dobrada a quantidade de Ulp e o tempo de clivagem foi reduzido para 1 hora a temperatura ambiente ($\sim 22^\circ\text{C}$). Após isso, era realizada centrifugação para remoção dos precipitados, seguida de uma etapa de cromatografia por exclusão molecular utilizando a coluna Superdex 75 10/300 GL (GE). Foi possível separar as populações da construção HsbD-Nt (Figura 55C).

As amostras de HsbD-Nt foram analisadas por espectrometria de massas, tanto pelo método de clivagem com tripsina quanto para o método de injeção direta, de acordo com os procedimentos descritos nas seções 2.5.2 e 2.5.3. As amostras preparadas pelo método por tripsinização não foram detectadas pelo equipamento de espectrometria de massas. Isso pode estar atribuído a massa pequena das proteínas e ao baixo rendimento obtido com a extração em gel. Com a amostra preparada pelo método de injeção direta, fragmentos correspondentes a HsbD-Nt foram detectados (dados não mostrados). Experimentos adicionais seriam necessários para concluir o que de fato leva a formação dessas duas populações proteicas. A amostra com maior massa, segundo gel de poliacrilamida, enumerada como pico 3 (Figura 55D) foi utilizada em experimentos posteriores. A Tabela 14 resume os progressos obtidos nos experimentos com HsbD-Nt, assim como seu rendimento médio de purificação.

Tabela 14 - Resumo de resultados para HsbD-Nt (pico 3 da gel filtração).

Construções	PCR	Clonagem	Expressão	Solubilidade	Purificação	Rendimento*
HsbD-Nt						0,5 $\pm$ 0,3

Legenda: Em azul estão as etapas concluídas sem dificuldades, em amarelo são as etapas apresentaram algum problema ou precisaram passar por otimização.

*Médias baseadas em três ou mais purificações independentes (medidas depois da exclusão molecular).

Fonte: Elaborada pelo autor.

3.1.4 Estudos de oligomerização

A mudança conformacional, acompanhada da mudança de estados de oligomerização, é uma das estratégias para regulação de proteínas. Isso confere importância a abordagem experimental na investigação da função. Construções de PA0847 apresentaram diferentes picos de eluição logo nas purificações (Figura 53), o que trouxe o indício que PA0847 poderia utilizar-se dessa estratégia, que será tratada nessa seção.

3.1.4.1 Análise da copurificação do c-di-GMP e sua influência sobre perfis de eluição

As diferentes populações em solução observadas para construções de PA0847 podem estar relacionadas com diferentes conformações da proteína e podem refletir o comportamento de PA0847 *in vivo*. Como foi observado para outras DGCs, diferenças populacionais foram consequência da ação do domínio GGDEF, que é copurificado com o c-di-GMP quando expressas sistemas heterólogos. Um exemplo é a WspR, com dímeros alongados quando ligada a c-di-GMP e com formas globulares na forma apo.¹⁰³ Foi observado a atividade de WspR a é dependente da formação de tetrâmeros.¹¹³ Outro exemplo é a tDGC, que forma monômeros e dímeros, com e sem o c-di-GMP ligado, respectivamente.¹¹² Algo semelhante visto nessas proteínas pode ocorrer para PA0847, dado que ela é predita ativa e possui o RXXD (Figura 47).

Analises por cromatografia em fase reversa das construções de PA0847, como descrita na seção 2.4.2, de fato indicaram a presença de c-di-GMP nas amostras purificadas de HPG-3 e PG (Figura 56).^v Portanto, para obter amostras sem o dinucleotídeo foi adaptado um protocolo de tratamento com fosfodiesterase (PDE). O tratamento consiste na inclusão de uma etapa de incubação, na cromatografia de afinidade e antes da clivagem com Ulp, com a PDE STM3615²⁴²⁻⁶⁶⁸ e o cofator da PDE, MgCl₂ (seção 2.3.2). A escolha foi devido ao fato que já havia comprovada a atividade da STM3615 *in vitro* em um trabalho anterior e se encontrava disponível em nosso laboratório.^{16,w} Apesar dos resultados da presença de c-di-GMP nas amostras purificadas serem sensíveis às condições de expressão e aos procedimentos de purificação, os resultados foram razoavelmente reproduzidos em diferentes lotes, preparados sob padronização na execução.

O tratamento com STM3615²⁴²⁻⁶⁶⁸, garantiu a completa remoção de c-di-GMP das construções HPG-3 e PG (Figura 56). Porém, esse resultado foi obtido em condições diferentes, pois para PG o tratamento foi mais vigoroso. Enquanto que uma incubação por 3h de STM3615²⁴²⁻⁶⁶⁸ na purificação de HPG-3 era suficiente para eliminar o sinal de c-di-GMP, para construção PG foi necessário utilizar o dobro de enzima, fazer o tratamento em solução e usar um tempo de incubação maior, de aproximadamente 16 h (seção 2.3.2).

^vAs amostras aplicadas ficaram em torno de 1 a 10 nmol. O menor padrão de c-di-GMP aplicado foi de 0,025 nmol (Figura 56). Portanto uma razão aproximada de 1 c-di-GMP para 100 de proteína já seria suficiente para detecção no equipamento.

^w Em testes preliminares desse trabalho, foi constatado por HPLC que a construção STM3615²⁴²⁻⁶⁶⁸ possui atividade (0,014±0,001) µM/s, utilizando 1 µM da enzima e 32 µM de c-di-GMP na reação, como descrito na seção 2.4.3.

Figura 56 - Cromatogramas do HPLC dos extratos das construções para análise de c-di-GMP copurificado.

Os cromatogramas mostrados são normalizados pela quantidade de proteína utilizada para análise, em nmol. As análises envolveram principalmente proteínas detentoras do GGDEF, conhecido por ligar o produto através do sítio inibitório RXXD. Os volumes de eluição das amostras foram comparados com aquele observado para o c-di-GMP padrão. Coluna utilizada: coluna Gemini 3 μ m C18 110 Å 150 x 4,6 mm.

Fonte: Elaborada pelo autor.

Para amostras de construções HPG-1, HPG-2 ou mesmo G não foi observado c-di-GMP copurificado. Para explicar essa ausência, pode-se levantar algumas hipóteses:

- Essas construções possuem menor afinidade pelo c-di-GMP, se comparado com HPG-3 ou PG, portanto, se houvesse c-di-GMP ligado na expressão, ele seria perdido durante as etapas de purificação e/ou na concentração da amostra antes da análise por HPLC;
- As formas ligadas ao c-di-GMP das construções que não possuem c-di-GMP ligado são instáveis e precipitam. Isso pode ser uma consequência em se trabalhar com proteínas truncadas ou mesmo em solução. Por isso as amostras finais de algumas das construções são apenas formas apo;
- As construções HPG-3 e PG poderiam apresentar maior atividade, dado que isso permitiria a observação da copurificação com o c-di-GMP. Isso incentivou ensaios de atividade.

As diferentes afinidades ou diferentes estabilidades apresentadas pelas construções com variantes do domínio HAMP pode ser, de fato, reflexo do papel desse domínio no mecanismo de regulação da porção citoplasmática de PA0847. Um exemplo para comparação é a ausência de interação com c-di-GMP reportada para construção HAMP e GGDEF de YfiN (YfiN^{HAMP-GGDEF}, resíduos 183-435). Entretanto, ao contrário da PA0847, YfiN não possui os resíduos necessários para interação com o dinucleotídeo, como a arginina no sítio Is(1) e outros (Figura 47). A ausência da ligação do c-di-GMP no domínio GGDEF de PA0847 reforça a importância dos domínios regulatórios para ligação ao produto em mecanismos de controle da atividade.

Para verificar se a copurificação de c-di-GMP alterava o perfil oligomérico das construções purificadas, foi realizado uma análise de SEC com e sem tratamento PDE (Figura 57).¹⁶ De fato, ocorre uma significativa mudança de perfil de eluição para as construções HPG-3 e PG. HPG-3 e PG sem tratamento apresentam dois picos sobrepostos de menor tamanho (que eluem por último). O pico extra que aparece pode ser resultado do c-di-GMP deslocando o equilíbrio para uma forma conformacional mais alongada, que por ser maior (maior raio hidrodinâmico) elui em um volume menor.

O efeito de diferenças das mudanças populacionais em função do tratamento fica mais evidente com a construção PG. Comparando os cromatogramas da proteína não tratada, com tratamento convencional (3 h a temperatura ambiente, 0,7 μ M de enzima na resina de afinidade) e com tratamento mais abrasivo (24 h na geladeira, 1,4 μ M de enzima e em solução) é possível observar uma diminuição gradual de uma populações e o aumento de outra (Figura 57). Com a construção HPG-3 foi obtida duas populações predominantes sem tratamento e uma população predominante com tratamento, provavelmente referentes a uma população ligada e outra apo.

Exceto quando não especificado, em todos os experimentos apresentados daqui em diante foram utilizados apenas o pico de eluição das proteínas correspondente a menor forma oligomérica (maior volume de eluição) resultantes da purificação por exclusão molecular e picos de maior forma oligomérica foram descartados. Isso foi necessário para conferir reprodutibilidade nos experimentos.

Figura 57 - Cromatogramas de purificações em Superdex 200 10/300 (GE) de construções de PA0847, com e sem tratamento PDE.

Amostras sem tratamento foram produzidas incubando a proteína na resina apenas com tampão (placebo). Pequenos deslocamentos dos volumes de eluição foram atribuídos a diferença de quantidade de proteína ou a outros fatores provindos dos aparatos experimentais. Para as proteínas HPG-3 e PG foram observadas variações nos volumes do número de picos de eluição, o que pode ser atribuído a presença de c-di-GMP nessas amostras; B) Para contornar diferenças experimentais, o cromatograma da exclusão molecular da amostra tratada por 3 h a temperatura ambiente ($T_a \sim 22^\circ\text{C}$) foi alinhada com aquele obtido com o tratamento por 16 h a 8°C (geladeira). Os cromatogramas originais estão na Figura 57A-PG. Para o alinhamento foram utilizados os picos de menor eluição (maior tamanho), destacado na figura. Nas análises por HPLC é possível observar que quanto mais abrasivo é o tratamento com a PDE, menor é a fração de c-di-GMP copurificado com a amostra. Concomitantemente, menor é a população oligomérica maior na cromatografia.

Fonte: Elaborada pelo autor.

3.1.4.2 Caracterizações das populações oligoméricas das construções de PA0847 por SEC analítica e SEC-MALS

Para análise de massas moleculares por exclusão molecular, de acordo com exposto na seção 2.6.1, o primeiro passo foi determinar a curva dos padrões globulares. Para o experimento, foi utilizada a coluna Superdex200 10/300 (GE), por resolver melhor os picos de eluição das construções (Figura 53B). O resultado dessa etapa de calibração é mostrado na figura a seguir (Figura 58).

Figura 58 - Curva padrão de massas para gel filtração, utilizando a coluna Superdex200 10/300 (GE).
Fonte: Elaborada pelo autor.

Foram utilizados para o experimento os picos de menor estado oligomérico das purificações, ou seja, de populações que eluem por último no experimento da gel filtração das purificações. As amostras de PA0847 foram preparadas de maneira idêntica à realizada para os padrões globulares (seção 2.6.1). Os cromatogramas resultantes estão resumidos na Figura 59. Dos resultados dessas cromatografias, foram calculados os volumes de eluição para os picos de eluição e com esses valores foram estimados os valores de massa molecular (Tabela 15).

Para de obter dados mais precisos e obter resultados para comparação, foi realizado outro experimento cromatografia de exclusão, mas acoplada a equipamento de detecção de espalhamento de luz (SEC-MALS). Amostras das construções, preparadas de maneira idêntica a análise por cromatografia analítica, foram analisadas por SEC-MALS, como descrito na seção 2.6.2. Desse experimento foi possível obter, dentre outros, a massa molecular das proteínas, independente da estrutura da mesma em solução (apêndice F). Um resumo dos cromatogramas obtidos é mostrado na Figura 60.

Figura 59 - Cromatogramas de exclusão molecular em Superdex200 10/300 (GE), usados para cálculo de massas.

Amostras de SDS-PAGE são exibidas ao lado de cada cromatograma e correspondem a amostras de picos equivalentes observados em purificações. BD: blue-dextran, polímero usado para medir o volume de eluição limite de resolução da coluna de exclusão molecular (V_0).

Fonte: Elaborada pelo autor.

Figura 60 - Cromatogramas produzidos em experimentos de SEC-MALS das construções de PA0847.

Linhas tracejadas representam o espalhamento de luz (LS), os traços contínuos são dos resultados da variação do índice de refração (dRI). Os cálculos das massas moleculares de cada pico são destacados em pontos vermelhos. As possíveis formas oligoméricas são mostradas nos tracejados, baseados nas massas esperadas das construções. *Coluna utilizada para as construções, exceto HPG-2: WTC-030N5 (Wyatt Technology); coluna utilizada para HPG-2: STR-C SEC-150 (Sepax).

Fonte: Elaborada pelo autor.

Para o domínio GGDEF isolado e na forma apo, era esperado monômeros. Entretanto foi observado um equilíbrio entre monômero e dímero para essa construção, o que é um dado novo para o domínio GGDEF. Um equilíbrio monômero e dímero também foi observado para as construções PG, HPG-2 e HPG-3 entretanto não apresentaram diferenças significativas nas proporções de dímeros, que se mantiveram iguais ou menores aproximadamente 10 % (Tabela 15). Baseado nisso, é possível supor que a interface dos dímeros observados é promovida principalmente pelo domínio G. Caso as construções tivessem uma propensão maior pela dimerização, como por exemplo pelas regiões flangeadoras clássicas do domínio PAS (N-cap e Jα) ou pela hélice AS2 do domínio HAMP, essa tendência iria se somar a do domínio GGDEF, o que elevaria a proporção de dímeros.

A construção P, que engloba apenas o domínio PAS de PA0847, se revelou exclusivamente monomérica. Isso poderia ser esperado uma vez que a construção não inclui as

regiões em hélice vizinhas, a hélice N-terminal (N-cap) nem a hélice C-terminal (J α). Essas hélices possuem função essencial na dimerização e transmissão da sinalização (seção 1.13.2 e 1.13.5). Existem exemplos de domínios PAS são capazes de se dimerizar pela folha central, mas esses exemplos são de eucariotos.

Tabela 15 - Resultados de oligomerização por SEC analítica (aSEC) e SEC-MALS.

Construção	População oligomérica (*)	MM calculada por aSEC (kDa) (**)	MM calculada por SEC-MALS (kDa) (***)	Rh por SEC-MALS (nm)	% em μ g por SEC-MALS
C4-1 (20,2 kDa)	Dímero (1)	46 $\pm$ 4	-	-	-
C4-2 (21,0 kDa)	Dímero (1)	50 $\pm$ 5	-	-	-
HPG-1 (38,2 kDa)	Tetrâmero (1)	210 $\pm$ 20	142,5 $\pm$ 0,4	5,5 $\pm$ 0,1	28,5
	Dímero (2)	70 $\pm$ 7	58,5 $\pm$ 0,9	3 $\pm$ 1	2,8
	Monômero (3)	37 $\pm$ 4	38,1 $\pm$ 0,2	3,05 $\pm$ 0,09	68,7
HPG-1^{AXXD} (38,2 kDa)	Tetrâmero (1)	-	-	-	-
	Dímero (2)	-	84 $\pm$ 2	18 $\pm$ 2	7,7
	Monômero (3)	-	39,4 $\pm$ 0,3	3,2 $\pm$ 4	92,3
HPG-2 (35,9 kDa)	Dímero (1)	100 $\pm$ 10	78 $\pm$ 3	5,9 $\pm$ 0,8	3,9
	Monômero (2)	44 $\pm$ 4	37,8 $\pm$ 0,2	3,12 $\pm$ 0,06	96,1
HPG-3 (34,6 kDa)	Dímero (1)	100 $\pm$ 10	67,7 $\pm$ 0,7	5,6 $\pm$ 0,4	7,2
	Monômero (2)	45 $\pm$ 4	34,2 $\pm$ 0,2	3,11 $\pm$ 0,05	92,8
HP (18,6 kDa)	Tetrâmero (1)	92 $\pm$ 9	66,7 $\pm$ 0,3	3,5 $\pm$ 0,1	17,3
	Monômero (2)	25 $\pm$ 2	22,3 $\pm$ 0,1	2,6 $\pm$ 0,1	82,7
P (11,2 kDa)	Monômero (1)	13 $\pm$ 1	11,5 $\pm$ 0,1	2,0 $\pm$ 0,2	100
PG (30,9 kDa)	Trímero (1)	240 $\pm$ 20	93 $\pm$ 2	3,6 $\pm$ 0,5	1,4
	Dímero (2)	82 $\pm$ 8	60,4 $\pm$ 0,4	3,6 $\pm$ 0,1	11,4
	Monômero (3)	32 $\pm$ 3	30,2 $\pm$ 0,2	2,87 $\pm$ 0,05	87,1
G (18,3 kDa)	Dímero (1)	35 $\pm$ 3	35,4 $\pm$ 0,3	2,9 $\pm$ 0,2	11,6
	Monômero (2)	17 $\pm$ 1	17,7 $\pm$ 0,1	2,14 $\pm$ 0,07	88,4

*Os números das populações oligoméricas fazem referência aqueles apresentados nas Figura 59 e Figura 60. A conclusão da forma oligomérica foi deduzida pela multiplicidade das massas obtidas com o SEC-MALS em relação as massas esperadas.

** aSEC: SEC analítica; os dados de volume de eluição foram coletados dos cromatogramas das construções e então esses dados foram convertidos em massa molecular através da curva padrão de proteínas globulares com massas moleculares conhecidas.

*** Dados calculados a partir dos dados de dRI e LS pelo programa ASTRA 7.0.1; os dados de raio de giro não puderam ser computados com precisão devido ao tamanho reduzido das partículas analisadas ($P(\theta) \approx 1$, para tamanho de partículas menores que $\lambda/20$, como exposto no apêndice F).

Fonte: Elaborada pelo autor.

Para as proteínas com o domínio HAMP era esperada a forma de dimérica ou monomérica (com dímeros transientes na forma ativa).¹¹¹ Entretanto, para construções com domínio HAMP inteiro, HPG-1 e HP, foram observadas populações tetramérica. Foi atribuído assim, a formação de tetrâmeros ao domínio HAMP. Apesar disso, diversos outros casos de proteínas com HAMP são descritos na forma dimérica, com estruturas resolvidas providas de HKs¹⁹⁰ e MCPs.¹³⁰ O domínio HAMP, por exemplo, precisa da forma dimérica para formação

do feixe de quatro hélices característico, descrito como o funcional. Um mecanismo de regulação promovido por um tetrâmero, assim, seria inédito para o domínio HAMP.

A construção HPG-1^{AXXD}, em relação a HPG-1, teve a população tetramérica com uma queda drástica e a população dimérica com um suave aumento. Isso sugere interferência do GGDEF, mais especificamente do domínio RXXD, na formação do tetramérica. Uma possível interpretação é que uma das interfaces diméricas, que compõe o tetrâmero, é dependente do GGDEF e por isso é quebrada com a mutação. Na segunda interface dimérica, que compõe o tetrâmero, deve ser promovida pelo domínio HAMP. Por isso HPG-1^{AXXD} se observa uma população maior de dímeros, se comparado com HPG-1.

A observação de monômeros das construções pode ser um artefato do trabalho com as proteínas parciais, sem o domínio receptor externo e regiões transmembrana, que possuem alta propensão para dimerização. Como visto nos resultados de exclusão analítica, construções periplasmáticas se apresentaram na forma de dímeros (Tabela 15), seguindo assim o padrão para receptores periplasmáticos. De acordo com outros exemplos de receptores de sinais externos descritos, a transmissão do sinal necessita de um dímero, que alterna a conformação com o ligante, por exemplo, de simétrica para assimétrica, ou vice-versa.

Uma construção da DGC transmembranar YfiN, contendo os domínios HAMP e GGDEF, também foi observada praticamente como monômero em outro trabalho. Para essa proteína, foi proposto um estado de transição dimérico transitório, necessário para explicar a atividade DGC. A mudança de resíduos importantes para formação do feixe do HAMP, tanto em YfiN quanto para PA0847, pode explicar a maior tendência dessas proteínas para formação de monômeros (Figura 61).¹¹¹

O alinhamento entre domínios HAMP mostra desvios na conservação do domínio em PA0847 e YfiN, o que também poderiam explicar a observação de dímeros não estáveis para essas proteínas (Figura 61). PA0847, por exemplo, possui uma prolina característica da hélice AS1 enquanto YfiN não possui o motivo DEXG da hélice AS2. Essa prolina e o motivo DEXG são conhecidos por formar uma rede no N-terminal do feixe de hélices do domínio HAMP, rede com função não bem estabelecida (Figura 29, seção 1.13.4). A ausência do motivo DEXG é característica de domínios HAMPs em série, o que não é o caso de YfiN. Além do mais, não é encontrada em YfiN e PA0847 a glicina no final da hélice AS1, que faz uma volta acentuada característica de outros HAMPs descritos.

Figura 61 - Alinhamento entre o HAMP de PA0847 e outros exemplares do domínio já estudados.

Na figura constam as sequências de duas proteínas usadas como modelo de diferentes enovelamentos do domínio HAMP (destacadas em linhas azuis). Uma delas é Af1503 é o modelo para empacotamento *coiled-coil x-da*, ou HAMP(A), e o HAMP2 de *P. aeruginosa*, modelo do empacotamento HAMP(B), no qual o final das hélices AS2 se aproximam de um *coiled-coil* dimérico. As DGCs PA0847 e Yfin estão destacadas em linhas vermelhas. Desvios dessas sequências para o padrão encontrado em HAMP(A) pode explicar desvios das formas oligoméricas observadas as DGCs. *O alinhamento foi gerado pelo programa HMMER, usando o HMM gerado pelo *seed* do Pfam para domínio.

Fonte: Elaborada pelo autor.

Apesar de HPG-2 e HPG-3 possuírem uma massa molecular menor, apresentam monômeros com volumes de eluição maior daqueles observados para HPG-1. Isso é evidenciado pelas massas calculadas para construções HPG-2 e HPG-3 em relação a HPG-1 no experimento de SEC analítica. As diferenças de eluição entre HPG-1 e HPG-2 ou HPG-3 foram reproduzidas em diferentes SECs, dentre elas, em análises de massa molecular com a coluna Superdex 75 10/300 (GE) (dados não mostrados).

Os raios hidrodinâmicos (rh), do experimento de SEC-MALS, concordaram com monômeros maiores para HPG-2 e HPG-3, em relação a HPG-1. Entretanto, utilizando o rh não é possível uma conclusão, frente ao tamanho próximo das partículas e os erros das medidas. O desvio para um volume maior para HPG-2 e HPG-3 em relação a HPG-1 pode ser indício de que essas proteínas estejam em diferentes conformações, o que pode vir a ter relevância fisiológica (na função da proteína *in vitro*) e talvez explicar diferentes atividades dessas construções (como será visto na seção 3.1.5). Para que essas conformações consigam explicar os resultados, é necessário postular que as conformações de HPG-2 e HPG-3 são mais alongadas que a apresentada por HPG-1.

O monômero da construção PG, assim como HPG-1, se comportou bem próximo a um monômero globular, com massa experimental próxima a esperada. Formas oligoméricas maiores da construção PG, como as triméricas, não foram reproduzidas durante os experimentos e representam uma pequena parcela do total medido. Por isso, outras formas oligoméricas da construção PG, além da monomérica e dimérica, poderiam ser um artefato da construção.

Para analisar qualitativamente o equilíbrio entre formas oligoméricas de amostras apo, foram utilizados como exemplos HPG-1 e PG. Os picos dos monômeros de HPG-1 e PG foram isolados de purificações para análises por SEC-MALS (Figura 62, Tabela 16). Essa análise é semelhante a apresentada na Figura 53A, com a diferença que aqui as amostras foram recromatografadas no SEC-MALS (Figura 62). Caso o equilíbrio fosse rápido, os picos deveriam fornecer as mesmas proporções de oligômeros.

Figura 62 - Análises de SEC-MALS dos picos oligoméricos da Superdex 200 10/300 (GE).

Os picos utilizados da purificação estão destacados em vermelho, no canto superior esquerdo de cada análise de SEC-MALS. Nas amostras utilizadas, não foi detectado c-di-GMP copurificado. (A) A tendência dos tetrâmeros de HPG-1 se dissociarem é maior que a tendência dos dímeros de PG se dissociarem (B). * Coluna utilizada para coluna HPG-1: STR-C SEC-150 (Sepax); coluna utilizada para PG: WTC 030S5 (Wyatt Technology).

Fonte: Elaborada pelo autor.

Para HPG-1, a recromatografia do tetrâmero da Superdex 200 10/300 (GE) recuperou 37,5 % de tetrâmeros no SEC-MALS. Por outro lado, utilizando monômeros da Superdex200 10/300 (GE) foram observados no SEC-MALS ≈ 15 % de tetrâmeros, menos que a metade do

resultado anterior. Para construção PG, uma população dimérica maior foi recuperada quando se aplicou no SEC-MALS os dímeros da Superdex 200 10/300 (GE) (Figura 62B). Por outro lado, a tendência em recuperar monômeros no SEC-MALS, partindo de dímeros, foi muito menor, de aproximadamente 10 % (Tabela 16).

Tabela 16 - Dados de SEC-MALS dos diferentes picos de eluição da purificação de exclusão das construções HPG-1 e PG

Construção	Oligomerização	Picos de maior oligomerização da Superdex 200 (Figura 62)		Picos referentes ao monômero da Superdex 200	
		Massa Molecular (kDa)	% em μ g	Massa Molecular (kDa)	% em μ g
HPG-1 (38,2 kDa)	Tetrâmero	175,4 $\pm$ 0,7	14,9	153,6 $\pm$ 0,2	37,5
	Monômero	40,5 $\pm$ 0,2	85,1	41,5 $\pm$ 0,2	62,5
PG (30,9 kDa)	Tetrâmero	123 $\pm$ 1	4,4	-	-
	Trímero	90,8 $\pm$ 0,6	11,7	-	-
	Dímero	59,8 $\pm$ 0,4	73,4	62 $\pm$ 1	4,4
	Monômero	31,4 $\pm$ 0,4	10,5	29,9 $\pm$ 0,2	95,6

*em destaque estão as mudanças nas populações de dímero e monômeros.

A coluna utilizada nesse experimento não possui resolução suficiente para separação de dímeros STR-C SEC-150 (Sepax). Dados computados através do programa ASTRA 7.0.1 (Wyatt Technology).

Fonte: Elaborada pelo autor.

Com os dados de recromatografia por SEC-MALS das populações obtidas em purificações, pode-se aferir que a conversão entre estados oligoméricos de PG é lenta. Isso indica que ambas as formas de PG são mais estáveis que as formas oligoméricas da construção HPG-1, que realiza a interconversão entre formas oligoméricas de maneira mais rápida que PG.

3.1.4.3 Impacto de c-di-GMP nos equilíbrios oligoméricos das construções de PA0847

Para analisar a interferência do c-di-GMP sobre o equilíbrio das populações das construções, foram feitas análises por SEC-MALS de amostras incubadas com c-di-GMP. Os cromatogramas e os respectivos resultados estão agrupados a seguir (Figura 63, Tabela 17). No experimento as amostras foram diluídas para 4 mg/mL com c-di-GMP, numa concentração 4 vezes maior que a da proteína, em molaridade. O excesso de c-di-GMP foi usado para favorecer um possível estado ligado das construções. As amostras foram reservadas a temperatura ambiente por 20 min ($\sim$ 22 °C), centrifugadas e submetidas ao experimento de SEC-MALS (Figura 63).

As amostras de HPG-1, PG e G das construções de PA0847 formaram precipitados quando incubadas com c-di-GMP. Isso comprometeu análises comparativas com a adição desse

ligante, não só nos experimentos de oligomerização, como também em outros experimentos de teste de interação entre as construções e o c-di-GMP.

Figura 63 - Resultados de SEC-MALS de construções da PA0847 com e sem c-di-GMP.

A diminuição dos sinais, extremo em alguns casos com em PG e G, foi atribuída a precipitação durante a incubação. Coluna utilizada: STR-C SEC-150 (Sepax).

Fonte: Elaborada pelo autor.

Tabela 17 - Resultados de SEC-MALS de comparação dos efeitos da incubação com c-di-GMP.

Construção	Oligomerização	Sem c-di-GMP		Com c-di-GMP	
		Massa (kDa)	% em µg	Massa (kDa)	% em µg
HPG-1 (38,2 kDa)	Tetrâmero	164±1	23,7*	153,7±0,5	8,4*
	Dímero	123,8±5	7,4*	83,1±0,2	25,4*
	Monômero	41±1	68,9	39,8±0,2	66,2
HPG-2 (35,9 kDa)	Dímero	80,4±0,5	5,6	79,8±0,9	4,2
	Monômero	38,1±0,2	94,4	38,5±0,2	95,9
HPG-3 (34,6 kDa)	Oligômero maior	5200±110	0,1	212±2	1,5
	Dímero	85±2	7,7	73,4±0,2	11
	Monômero	36,5±0,3	92,3	35,5±0,2	87,5
PG (30,9 kDa)	Oligômero maior	110±3	1,5	138±5	4,4
	Dímero	62,3±0,4	11,7	78±2	10,8
	Monômero	30,8±0,2	86,8	34,7±0,4	84,7
G (18,3 kDa)	Dímero	43,4±0,7	5,9	82±10	2,5
	Monômero	19,6±0,2	94,1	22,4±0,5	97,5

* Os dados de dímeros e tetrâmeros foram destacados pois foram uma das principais mudanças observadas com a adição de c-di-GMP. As proporções dessas populações praticamente ficaram invertidas com a adição de c-di-GMP. Dados computados através do programa ASTRA 7.0.1 (Wyatt Technology).

Fonte: Elaborada pelo autor.

Apesar da precipitação, observa-se para HPG-1 que a população de tetrâmeros diminuiu muito e a de dímeros tem um ganho com a adição de c-di-GMP. Devido a melhora do sinal dos dímeros, a massa experimental fica bem próxima da esperada para dímeros, com ≈ 75 kDa (Tabela 17).^x O cromatograma de HPG-1 com c-di-GMP é semelhante ao obtido para a construção mutante HPG-1^{AXXD} (Figura 60). Como é sabido que o c-di-GMP pode se ligar ao sítio inibitório^{88,111}, é possível concluir desses resultados que a formação de tetrâmeros depende, além do domínio HAMP, do sítio inibitório RXXD.

3.1.4.4 Impacto da força iônica da solução tampão nos equilíbrios oligoméricos das construções de PA0847

Conhecendo as diferentes formas oligoméricas das construções de PA0847, seguiu-se os estudos de como essas populações poderiam estar correlacionadas com a função DGC. Uma maneira encontrada para alterar o equilíbrio oligomérico artificialmente foi alterando a concentração de NaCl da solução. Para verificar esse fato, foi utilizado o SEC-MALS. Nos experimentos, a amostra foi diluída de maneira que apenas as concentrações de NaCl se alterassem para 600mM, 300mM e 150mM. Depois da diluição, as amostras foram reservadas por aproximadamente 1 hora a temperatura ambiente (~ 22 °C). No tampão utilizado para aplicar as amostras, além do tris 25 mM com pH 8,0, foi fixado 600mM de NaCl. Isso porque ao variar a concentração de NaCl da fase móvel no SEC-MALS foi observado um deslocamento nos tempos de eluição das construções. Isso dificultou a sobreposição dos cromatogramas e assim as análises comparativas (dados não mostrados). Os resultados foram compilados na Figura 64 e as análises quantitativas foram organizadas na Tabela 18.

Foi observada uma queda das populações de tetrâmeros de HPG-1 e HP, conforme o aumento de concentração de NaCl. No experimento não foi observado alteração significativa nas populações diméricas, observadas para outras proteínas com domínio GGDEF. Mas foram observados também uma redução significativa da população das formas tetraméricas com a mudança do tampão de tris para hepes (dados não mostrados). Com isso, foi demonstrado que se pode variar principalmente as populações de tetrâmeros e de maneira desprezível para dímeros, alterando a força iônica ou o agente tamponante da solução (como hepes, dados não mostrados).

^x O desvio de valores experimentais para valores esperados foi comum em experimentos para eluições com sinal baixo ou com sinais sobrepostos.

Figura 64 - Sobreposição dos perfis de eluição no experimento de SEC-MALS de construções da PA0847 com diferentes concentrações de NaCl na amostra aplicada. Coluna utilizada: STR-C SEC-150 (Sepax). Fonte: Elaborada pelo autor.

Tabela 18 - Variação das populações oligoméricas frente a concentração de NaCl na amostra aplicada.

Construção	Oligomerização	Média das massas moleculares (kDa)	150 mM	NaCl: 300 mM (% em µg)	600 mM
HPG-1 (38,2 kDa)	Tetrâmero	150±8	17,9*	14,2*	8,5*
	Monômero	38,2±0,8	82,1	85,8	91,5
HPG-2 (35,9 kDa)	Dímero	80±30	2,4	1,8	3,3
	Monômero	35,7±0,6	97,6	98,2	96,7
HPG-3 (34,6 kDa)	Dímero	71±6	9,8	8,3	9,3
	Monômero	34,6±0,2	90,2	91,7	90,7
HP (18,6 kDa)	HMWO**	190±10	5,6	5,9	5,9
	Tetrâmero	81±4	14,6*	12,1*	9,1*
	Monômero	21,2±0,3	79,8	82,0	85,0
PG (30,9 kDa)	Dímero	67,2±4	9,1	9,9	9,9
	Monômero	31,3±0,4	90,9	90,1	90,1
G (18,3 kDa)	Dímero	33±8	0,8	1,9	1,3
	Monômero	18,1±0,3	99,2	98,1	98,7

Dados computados através do programa ASTRA 7.0.1 (Wyatt Technology).

*em destaque estão a mudanças mais significativas, na população das construções com tetrâmeros;

**HMWO: oligômeros de alta ordem.

Fonte: Elaborada pelo autor.

3.1.5 Estudos da atividade catalítica das construções de PA0847 contendo o domínio GGDEF

Os ensaios de atividade foram aplicados para verificar se as construções de PA0847 apresentariam diferentes taxas de conversão de c-di-GMP. Essas diferenças poderiam

fundamentar hipóteses sobre como a função e dinâmica dos domínios. Os primeiros testes verificaram a atividade das construções utilizando meio reacionais analisados por cromatografia de fase reversa, como exposto na seção 2.4.4. São mostrados, a seguir, o gradiente com as curvas padrões de GTP e c-di-GMP (Figura 65 A), como também as curvas das reações (Figura 65 B).

Figura 65 - Análises de HPLC de padrões e de amostras de reações com GTP (A) e de reações com construções da PA0847 (B).

Os cromatogramas das reações mostrados são referentes a concentrações de 10 µM e de amostras deixadas incubadas ON (≈ 16 horas). Essas curvas foram realizadas com amostras purificadas com protocolo sem tratamento com fosfodiesterase, por isso parte do sinal pode ser provindo de c-di-GMP copurificado.

Fonte: Elaborada pelo autor.

Houve a ausência de atividade apenas para o domínio GGDEF isolado, mesmo que esse seja capaz de formar dímeros, oligômero necessário para atividade. Isso mostra a importância dos domínios regulatórios na conformação correta para a atividade.⁹⁶ A presença do domínio PAS foi suficiente para o surgimento da atividade, o que demonstrou a importância desse domínio regulatório para a atividade.

Uma vez verificadas as atividades por HPLC, foram realizados outros testes com o *kit* comercial EnzChek (Invitrogen) para otimização da concentração de enzima (seção 2.7). Esse *kit* foi escolhido por já é utilizado para caracterização de DGC.¹¹³ Foi constatado dos testes preliminares que concentrações de 1 a 2 µM das construções seriam adequadas para os experimentos (resultados não mostrados). A escolha dessas concentrações levou em consideração que as construções apresentaram taxas de conversão de c-di-GMP distintas e buscou-se utilizar uma única para facilitar comparações entre as construções.

Histogramas com os resultados de atividade foram elaborados e estão mostrados a seguir (Figura 66).^y A ordem de grandeza dos resultados das construções de PA0847 ficou próximo

^y Foi necessário usar para os ensaios como temperatura padrão 30 °C, dada a limitação do leitor de placas disponível Spectramax® i3 (Molecular Devices). Isso pode ter contribuído para baixa atividade das construções.

do esperado para uma DGC, com k_{cat} na ordem de 1 min^{-1} e $0,1 \text{ min}^{-1}$ (conversões por minuto) o que equivale a $\approx 0,02 \text{ s}^{-1}$ e $\approx 0,002 \text{ s}^{-1}$ (conversões por segundo).⁸⁸

Figura 66 - Resultados de cinética das construções de PA0847, com e sem NaCl adicionado.

O fator de correção obtido utilizando as curvas com os controles positivos de pirofosfato foi de $(2,5 \pm 0,2)$.

Fonte: Elaborada pelo autor.

Nos ensaios, a maior atividade foi apresentada pelo domínio HPG-1. A baixa atividade de PG talvez esteja relacionada com uma maior afinidade pelo c-di-GMP (maior inibição),

como indicado pela maior tendência da construção de ser copurificada com c-di-GMP (seção 3.1.4.1). Uma vez que HPG-1 é a única que possui o domínio GGDEF que é capaz de formar tetrâmeros, buscou-se elucidar se esses tetrâmeros eram os responsáveis pela elevada atividade de HPG-1. Por isso, foram monitoradas também reações nas quais foram acrescentados 600 mM de NaCl (Figura 66). As taxas de conversão de todas as reações monitoradas com a adição de NaCl sofreram uma queda substancial, inclusive, de maneira inesperada, para as enzimas do *kit* usado para fazer o monitoramento, como indicado pelo controle com pirofosfato (PPi, Figura 66A).

Foi necessário assumir, para permitir comparações, que o efeito do aumento da força iônica para as enzimas do *kit* comercial seja próximo aquele observado para as construções de PA0847. Partindo disso, foi calculado um fator de correção, igual a divisão entre as taxas de conversão de reações com PPi, sem e com excesso de NaCl. Então, os resultados de todas as reações monitoradas com 600 mM de NaCl foi multiplicado pelo fator de correção estipulado ($2,5 \pm 0,2$) (Figura 66B).

Mesmo compensando o efeito da mudança da força iônica, através de um fator baseado no padrão de PPi, é possível observar que a taxa de conversão da construção de HPG-1 ainda é bem reduzida. Essa redução foi de aproximadamente 50%, com a adição de NaCl ($\approx 12 \cdot 10^{-3} \mu\text{M/s} \rightarrow \approx 6 \cdot 10^{-3} \mu\text{M/s}$, Figura 66B). Recuperando os dados de SEC-MALS (Tabela 18), os tetrâmeros obtidos com 150 mM e 600 mM de NaCl são reduzidos também por volta de 50% ($\approx 18\% \rightarrow \approx 9\%$), acompanhando a queda observada na atividade. Isso indica uma relação entre a formação de tetrâmeros e a atividade DGC para construção HPG-1. Outra hipótese seria que a conformação que permite uma maior atividade apresentada por HPG-1 é a mesma que permite a formação de tetrâmeros, baseados nos resultados obtidos.

Para as outras construções que apresentam dímeros, de acordo com a concentração de NaCl, não ocorre uma mudança significativa na porcentagem de dímeros (Tabela 18), nem da atividade DGC (Figura 66B). Isso é contundente com a não alteração significativa observada da população de dímeros com a variação de NaCl na solução das amostras (Figura 64, Tabela 18). A construção da proteína HPG-1^{AXXD}, construção com população praticamente não detectável de tetrâmeros, a diferença entre a atividade sem NaCl e a atividade com NaCl corrigida é de $\approx 15\%$, não tão expressiva quanto a observada para HPG-1.

Foram realizadas ainda curvas de cinética e inibição por c-di-GMP para HPG-1 e HPG-1^{AXXD}, como apresentado na seção 2.7. Pelas curvas de Michaelis-Menten e de inibição utilizando HPG-1 foi possível constatar a baixa afinidade de HPG-1 pelo substrato, com $K_m \approx 50 \mu\text{M}$ e pelo produto, com $K_i \approx 10 \mu\text{M}$ (Figura 67). A mutação sítio dirigida no domínio

regulatório promoveu um aumento significativo, de aproximadamente 7 vezes em $V_{\text{máx.}}$, $\approx 0,07 \mu\text{M/s}$, se comparado ao obtido com a construção selvagem, com $\approx 0,01 \mu\text{M/s}$. É necessário lembrar que os valores obtidos de K_m , k_{cat} e K_i são subestimados, dado que a proteína precipita com ligante e o inibidor é o próprio produto, ou seja, a concentração de c-di-GMP é maior que utilizada para cálculos pois o c-di-GMP é produzido pela enzima.

Figura 67 - Gráficos de atividade das construções de PA0847, com e sem a adição de NaCl no meio reacional. Fonte: Elaborada pelo autor.

Para o mutante, a taxa de conversão é equivalente aproximadamente 4 min^{-1} . A constante de inibição K_i para HPG-1 está condizente com o valor fisiológico do c-di-GMP intracelular bacteriano, estimado em até $10 \mu\text{M}$.⁸² Foi possível constatar que a construção HPG-1^{AXXD} não é inibida pelo substrato, dada a faixa de concentrações de c-di-GMP testada. Isso leva a afirmação de que o sítio RXXD é o principal fator inibitório da proteína PA0847.

3.3.6.2 Cristalização e difração de raios X das construções da PA0847

Estudos para obtenção da estrutura cristalográfica são fundamentais em biologia molecular, pois auxiliam na formulação, ou mesmo na comprovação, de hipóteses sobre a função. Como primeiro passo, foram realizadas triagens por tampões para encontrar algum que forneça uma amostra mais homogênea em solução, o que poderia elevar as chances de cristalização. A homogeneidade das amostras foi monitorada por DLS (seção 2.11.1). As construções C4-1 e C4-2 tiveram os valores de polidispersividade reduzidos com o uso tampão CHES pH 9,5 e 500mM de NaCl. Resultados de DLS com as construções citoplasmáticas HPG-2 e PG ficaram com a polidispersividade reduzida (dos monômeros) em tampão HEPES pH 7,5

contendo 500mM de NaCl (dados não mostrados). Para as construções HPG-1 e HPG-3 experimentos de otimização de tampão não foram realizados e o tampão assumido foi o mesmo das purificações.

As construções C4-1, C4-2, HPG-1, HPG-2, HPG-3 e PG foram submetidas a triagens de condições de cristalização (seção 2.11.2). As concentrações testadas foram 30, 20 e 10 mg/mL, exceto para construção HPG-1 (15, 10 e 5 mg/mL) e HPG-1^{AXXD} (7 mg/mL) devido ao menor rendimento dessas construções. Construções PG e HPG-3 também tiveram amostras testadas com e sem tratamento PDE.

Para diversas condições das construções citoplasmáticas HPG-1, HPG-2, HPG-3 e PG foram observadas a formação de esferolitas. Isso pode ser efeito da não homogeneidade das amostras que, como visto, se apresentam em equilíbrio oligomérico em solução. Por outro lado, foram obtidos cristais para construção periplasmática C4-2, em duas condições do *kit* PEGION (Hampton Research). A primeira condição foi 0,2 M acetato de magnésio, 20 % PEG3350, pH 7,9 (Figura 68A, condição 1). A segunda condição foi formato de magnésio 0,2 M, 20 % PEG 3350 (m/vol), com pH 6,0 (Figura 68A, condição 2). Em nenhuma condição foi observado cristais para C4-1. Isso reforçou a importância da estratégia de trabalho com diferentes construções, mesmo que com tamanhos próximos.

Foi dada continuação com o refinamento da condição de cristalização com formato, uma vez que resultaram em cristais maiores na triagem inicial. Foram testadas as concentrações de formato a 0,2, 0,15 e 0,1 e 0,5 M. Nas otimizações, as concentrações de proteína foram de 20, 15 e 10 mg/mL. Para concentração de PEG foram 20, 21, 22, 23, 24, 25, 26, 27, 28 e 29 (% m/vol). Os melhores cristais surgiram nas condições de 0,1 M de formato com 28 % de PEG (Figura 68B). Também foram obtidos cristais de C4-2 com a selenometionina incorporada. Com os mesmos passos de otimização, foram obtidos cristais maiores com 22 % de PEG e concentração de 0,1M de formato.

Cristais foram submetidos a difração de raios-X nos dias 15/08/14, 03/12/14 , 05/09/15 no LNLS, Campinas- SP, usando algumas técnicas pré-coleta como *annealing*, desidratação e crioproteção do cristal.²⁴⁷ Com os melhores resultados de coleta do dia 05/09/15, com cristais com selenometionina incorporados, foi possível realizar o processamento e obter algumas estatísticas (Tabela 19).

Tabela 19 - Dados de coleta e de processamento para C4-2 com selenometionina (LNLS).

Resolução (Å)	3,7 – 3,5	Parâmetros de rede:	
Grupo espacial	P3 ₁ 21 ou P3 ₂ 21	a, b, c (Å)	84,5, 84,5, 137,9
Redundância	2,9 (2,2)	α , β , γ (°)	90, 90, 120
Completeza (%)	94,3 (85,6)	Conteúdo de solvente	62 %
I/ σ	5,7 (2,8)	Reflexões totais	20811 (2329)
λ (Å)	0,9789	Reflexões únicas	7123 (1016)
Proteínas por unidade assimétrica	2	R _{sym}	19,4 (40,9)

Fonte: Elaborada pelo autor.

Figura 68 - Cristais da construção periplasmática da PA0847 C4-2. A) Condições iniciais que forneceram cristais na primeira triagem. A figura mostrada é de um cristal obtido na condição 1 (acetato de magnésio) com 10mg/mL de proteína, e os cristais da foto marcada como condição 2 (formato) são daqueles obtidos com 30 mg/mL. Os valores em parênteses são referentes aqueles da camada externa dos dados de reflexão; B) Cristais otimizados com selenometioninas incorporados à proteína. A condição utilizada foi 0,1 M de formato de magnésio e 22% de PEG. C) Padrão de difração obtido com os cristais com selenometionina. A faixa de maior resolução é de 3,5 Å; D) Cristais de C4-2 incorporados com selenometionina obtidos, acrescentando o detergente n-dodecil- β -maltosídeo. Esses cristais foram obtidos nas mesmas soluções empregadas com amostras com selenometionina, entretanto com 15 mg/mL (figura) e 7,5 mg/mL; E) Os cristais sem o detergente n-dodecil- β -maltosídeo se formam no período de 1 dia mas se desfazem (morrem) com o tempo. Com 15 dias o resultado é mais evidente. Já com o detergente os cristais se formam em dois ou três dias e não se desfazem com o tempo.

Fonte: Elaborada pelo autor.

Entretanto, esses dados não foram suficientes para resolução da estrutura. Isso foi atribuído a relativa baixa resolução (em torno de 3,5 Å, Figura 68C) e baixa incorporação de selenometionina (devido ao baixo sinal anômalo observado). A baixa qualidade da coleta foi atribuída também ao fato dos cristais cresceram sobre o precipitado, o que faz com que sua manipulação seja dificultada. A baixa qualidade dos cristais pode estar relacionada também a

baixa instabilidade dos mesmos, dado que eles se desfazem em um período de aproximadamente 15 dias (Figura 68E).

Visando elevar a qualidade dos cristais, foi realizada uma triagem adicional por aditivos, usando o *kit* Additive Screen™ (Hampton Research), como descrito na seção 2.11.2. Em quase todas as condições os cristais de C4-2 se apresentaram de maneira semelhante aos já obtidos, exceto para a condição com o detergente n-dodecil- β -maltosídeo (5 % m/vol), que diminuiu a quantidade de precipitado da proteína e forneceu cristais com dimensões mais próximas a cubos (Figura 68D).

Os cristais com n-dodecil- β -maltosídeo foram submetidos a experimentos de difração no síncrotron *Diamond Light Source*, em Oxford, Inglaterra. Isso ocorreu em abril de 2015. Esses experimentos foram conduzidos pelos doutores Bruno Matsuyama e Flávio Rosseto. Apesar dessa mudança em aspectos do cristal com o uso do detergente, foi obtida a mesma qualidade na difração.

3.1.6 Hipóteses da regulação de PA0847 pelo domínio periplasmático (CHASE4)

Uma vez que não foi obtida a estrutura cristalográfica do domínio periplasmático de PA0847, recorreu-se a modelos de predição para continuação das análises. A estrutura do CHASE4 não está disponível, mas trabalhos já é citado o enovelamento desse domínio como PDC. Submetendo a região predita periplasmática de PA0847⁶²⁻³¹⁵ ao programa de predição de estrutural I-TASSER²¹⁴ foram obtidos dois domínios PDC (*double-PDC*)(Tabela 20).

Tabela 20 - Resumo de estruturas utilizadas para construção do modelo da porção periplasmática de PA0847 (resíduos 62-315).

Colocação	PDB	Ident. 1	Ident. 2	Cobertura	Z-score	Proteína	Ligante
1	4xmq	0,09	0,18	0,89	0,94	Tlp3 (CcmL)	Sulfato
2	3c30	0,10	0,20	0,80	1,41	LuxQ	LuxP*
3	3lid	0,11	0,18	0,94	1,26	VpHK1S-Z8**	Fosfato
4	1ojg	0,16	0,13	0,53	1,29	DcuS	D-malato
5	2hj9	0,11	0,18	0,77	1,03	LuxQ	LuxP
6	3lif	0,08	0,23	0,88	1,22	RpHK1S-Z16	-
7	1p0z	0,13	0,15	0,40	1,41	CitA	Citrato
8	3e4p	0,12	0,22	0,85	1,29	DctB	Ác. malônico
9	5ere	0,11	0,17	0,81	0,81	DSM5692	Citosina
10	2p7j	0,11	0,19	0,95	1,79	VpHK1S-Z8**	Sulfato

* identidade 1: identidade, no alinhamento, entre a parte da sequência do molde usado e sequência submetida para predição; identidade 2: identidade, no alinhamento, entre toda sequência do molde usado e a sequência submetida para predição; cobertura: número de resíduos do molde alinhados dividido pelo número de resíduos da sequência submetida para predição; Z-score normalizado: quando acima de 1 indica um alinhamento confiável e que provavelmente a proteína molde possui o mesmo enovelamento da proteína submetida para predição.^{214, 248}

Fonte: Elaborada pelo autor.

O enovelamento PDC também foi observado em LuxQ (Figura 69A), DctB e algumas outras estruturas resolvidas por Zhang e Hendrickson (Figura 69B).¹⁴⁶

Figura 69 - Mecanismos propostos para proteínas com enovelamento *double*-PDC, também identificado para porção periplasmática de PA0847. A) O complexo Lux-Q-LuxP é constitutivo, ou seja, é formado sem a necessidade do ligante. LuxQ possui a atividade quinase na ausência do autoindutor 2 (AI-2, ou S-THMF-borato²⁵). Na presença de AI-2, rearranjos estruturais envolvendo principalmente o loop FG e o lobo entre os domínios PDC, fazem LuxQ assumir a conformação fosfatase; B) Estruturas diméricas de sensores, sem SBPs associados, de receptores com dois PDCs, com e sem ligantes, mostraram mudanças estruturais que podem explicar os mecanismos de transmissão da sinalização pela membrana dessas proteínas. Duas dessas estruturas mostraram um deslocamento entre as hélices dos limítrofes da estrutura, paralela a membrana (PDB: 3li9). Em outro exemplo, a presença do ligante promove a curvatura assimétrica da região sensora (3lid); C) Estruturas do modelo do I-TASSER para a porção periplasmática da PA0847 (aminoácidos 62-315), com destaque para região classificada como CHASE4 (80-249). Estruturas semelhantes da região periplasmática de LuxQ e de um sensor denominado vpHK1, ilustrando a semelhança estrutural entre essas proteínas que possuem dois domínios PDC.

Fonte: A) Adaptada de NEIDITCH et al¹⁵⁶, B) Adaptada de ZHANG; HENDRICKSON¹⁴⁶, C) Elaborada pelo autor.

A predição teve pontuação aceitável para topologia global, que deve ser maior que 0,5,²⁴⁸ sendo que o *TM-score* obtido foi de 0,52. Já para o parâmetro C-score, que estima a qualidade do modelo, geralmente é da faixa [-5,2], para o modelo da região periplasmática de PA0847 foi obtido -1,55, o que indica um modelo razoável.

Pela análise do resultado de estrutura predita é possível constatar que a parte periplasmática que engloba o CHASE4 corresponde a porção PDC membrano distal. A estrutura melhor colocada na construção do modelo pelo I-TASSER da região periplasmática de PA0847, de PDB 4xmj, possui um íon sulfato ligado (20). Outros dois PDBs, 2p7j e 3lid, utilizados para montar o modelo estrutural, são ambos referentes mesma proteína, de código Uniprot Q87SR8. A primeira dessas estruturas foi resolvida com sulfato e outra com fosfato, semelhantes quimicamente. Outras estruturas utilizadas como modelo vem de LuxQ, apo e ligada (Tabela 20). O C-score para as predições do ligante, que deve estar entre [0,1] foram bem baixas, sendo a melhor para d-malato, ligante provindo da estrutura de DcuS (PDB: 3by8), com o *C-score* de 0,22.

Para tentar aferir o papel da regulação do domínio periplasmático CHASE4, foi considerado que PA0847 (genoma de PAO1: 925007-927214), que corresponde a PA14_53310 (de PA14, genoma: 4724553-4726598) é predita participar de um operon com PA0846 (924182-924922; PA14_53330 do genoma PA14: 4726683-4727423)(Figura 70).^{204,249}

Figura 70 - Predição de operon de PA0847 e PA0846, disponibilizado pelo servidor pseudomonas.com.

Os genes de PA0847 e PA0846 da cepa PAO1 corresponde a cepa PA14 aos genes PA14_53310 e PA53310. PA0846 é anotada como transportador de sulfato CysZ, devido a elevada similaridade com aquela descrita em *E. coli*. A imagem é disponível no modo de visualização GBrowse (*The Generic Genome Browser*). As predições através do DOOR (*Database for prOkaryotic OpeRons*), disponibilizados também pelo servidor pseudomonas.com; B) Esquematização resumida da via de absorção do sulfato em *E. coli*. APS - adenosina 5'-fosfosulfato.

Fonte: A) Adaptada de PSEUDOMONAS...²⁰⁴, B) Adaptada de ZHANG et al.²⁵⁰

PA0846 possui aproximadamente 40 % de identidade com CysZ, um transportador de sulfato de *E. coli*.²⁵⁰ Esse transportador é composto essencialmente por hélices embebidas na

membrana e possui estrutura cristalográfica (PDB: 3TX3), com trabalho a ser publicado (Figura 70B).

Dado esses resultados, foi considerado a possibilidade de que PA0847 participe da rota desse íon ou mesmo por reconhecimento do mesmo. Embora nem PA0847 e PA0846 possuam níveis de transcrição alterados de maneira significativa segundo trabalho com *P. aeruginosa* em meios suplementados com sulfato ou ausência do mesmo²⁵¹, essas proteínas não devem sofrer alteração nos níveis de transcrição mas sejam iniciadores de cascatas de sinalização. Testes por calorimetria de titulação isotérmica (ITC) não indicaram a interação com o íon utilizando a construção C4-2 (dados não mostrados). Isso sugere que deve haver uma PBP auxiliando na ligação ou que o segundo PDC, membrano proximal, seja importante para o reconhecimento do íon, caso ele seja mesmo o ligante da porção periplasmática de PA0847.

3.1.7 Prováveis funções estruturais do domínio PAS

Outra questão do trabalho foi buscar determinar a função do domínio PAS em PA0847. Uma das possibilidades seria a interação com alguma outra molécula¹⁵⁴, integrando uma via intracelular com a sinalização recebida pelo domínio periplasmático. Uma segunda hipótese seria o auxílio na ligação ao c-di-GMP, como visto em PleD⁹⁴⁻⁹⁵ e em outros exemplos¹¹¹, na qual o c-di-GMP está na interface intradomínios (seção 1.8.2, Figura 14). Uma terceira hipótese seria o domínio possuir apenas função estrutural.^{119,164-165} Se houvesse a participação da regulação pelo HAMP isso não seria algo novo, pois estudos já mostraram a regulação do PAS da proteína quimiotática Aer de *E. coli* sobre a hélice AS2 do domínio HAMP da mesma cadeia polipeptídica.¹⁷⁸⁻¹⁷⁹

Para buscar indícios de algum ligante do PAS, foi feita a predição da estrutura do domínio PAS de PA0847 com o I-TASSER.²¹⁴ O modelo construído utilizou algumas proteínas que se ligam a FAD e uma que se liga FMN (Tabela 21). O modelo foi considerado confiável, dado os bons resultados dos parâmetros resultantes da predição. Para o modelo foi obtido o valor de 0,94 para *C-score* (parâmetro de qualidade do modelo que varia entre -5 a 2). Outro parâmetro, *TM-score* foi obtido o valor de 0,84±0,08. Esse parâmetro precisa ser maior que 0,5, limiar indicativo de uma topologia correta da estrutura predita.²¹⁴ O *C-score* referente ao ligante foi de 0,88 (nesse caso varia de 0 a 1) para o ligante FMN, baseado na estrutura LOV2 de uma fototropina de *Arabidopsis thaliana* (PDB: 4eep). Isso reforçou que o ligante poderia ser alguma das moléculas com grupo flavina.

Tabela 21 - Resumo de estruturas utilizadas para construção do modelo da porção PAS de PA0847 (resíduos 402-504)

Colocação	PDB	Ident. 1	Ident. 2	Cobertura	Z-score	Proteína	Ligante
1	3ewk	0,23	0,31	0,99	2,82	MmoS	FAD
2	3lyx	0,25	0,26	1	2,49	CPS_1291	-
3	3ewk	0,23	0,26	0,99	4,24	MmoS	FAD
4	1x0o	0,13	0,16	1	1,59	ARNT (PAS-B)	-
5	1x0o	0,13	0,16	1	1,04	ARNT (PAS-B)	-
6	3lyxA	0,25	0,26	1	4,04	CPS_1291	-
7	1x0o	0,13	0,16	0,99	1,62	ARNT (PAS-B)	-
8	2gj3	0,23	0,27	1	2,19	NifL	FAD
9	5iu1	0,18	0,22	1	3,95	PpANR MAP3K	-
10	1n9l	0,2	0,22	1	2,25	Phot-LOV1	FMN

* identidade 1: identidade, no alinhamento, entre a parte da sequência do molde usado e sequência submetida para predição; identidade 2: identidade, no alinhamento, entre toda sequência do molde usado e a sequência submetida para predição; cobertura: número de resíduos do molde alinhados dividido pelo número de resíduos sequência submetida para predição; Z-score normalizado: quando acima de 1 indica um alinhamento confiável e que provavelmente a proteína molde possui o mesmo enovelamento da proteína submetida para predição.^{214,248}
 Fonte: Elaborada pelo autor.

Para confirmar um potencial ligante, foram realizadas também análises de alinhamento, utilizando diversas sequências de PAS já caracterizados (Figura 71A). Com o alinhamento foi montada uma árvore filogenética com o programa MEGA. Foi possível observar que a PA0847 ficou sobre próximo aos domínios que se ligam a FMN e FAD. Entretanto, nenhum dos resíduos característicos para a ligação dessas moléculas foi encontrado no alinhamento. A ausência de picos de absorbância característicos do FAD, nas construções de PA0847 que possuem o domínio PAS (dados não mostrados), suportam a hipótese que o PAS de PA0847 não contenha esses cofatores.

Na árvore filogenética, o domínio PAS de PA0847 fica bem próximo do mesmo domínio de RbdA. Portanto, assim como em RbdA, o domínio PAS de PA0847 deve possuir alguma função estrutural. Como as hélices N-cap e J α , flangeadoras do domínio, não promoveram um ganho na população dimérica (Figura 66), a função do domínio PAS deve ser outra. Como visto, a construção PG apresentou baixa atividade (Figura 72), c-di-GMP copurificado (Figura 62) e amostras copurificadas com nucleotídeo estáveis, ao contrário da amostra quando incubada com c-di-GMP (Figura 63).

Levando em consideração que o domínio GGDEF isolado não apresenta atividade (Figura 72), coletivamente esses resultados e análises indicam que o domínio PAS de PA0847 pode ainda auxiliar o correto posicionamento dos domínios GGDEF para a atividade e estabilizar uma conformação inibida, na qual o c-di-GMP está ligado na interface PAS-GGDEF intercadeia (forma inibida 2). Desses resultados, também não pode ser excluída a possibilidade que PAS estabilize uma conformação intracadeia GGDEF-GGDEF (forma inibida 1), o que explica a instabilidade das proteínas quando se adiciona c-di-GMP.

Figura 71 - Parte do alinhamento de alguns domínios PAS conhecidos (B) e árvore filogenética a construída a partir do mesmo (B).

A) Algumas sequências de domínios PAS já bem caracterizados foram selecionadas manualmente e alinhadas com o programa ClustalW. Em destaque no alinhamento estão alguns resíduos clássicos para coordenação de FMN e FAD¹⁵⁴, os quais a sequência de PA0847 não apresenta. Em parênteses estão a região predita como PAS e que abrange o motivo conservado DIT; B) Com o alinhamento, foi construída uma árvore filogenética utilizando o programa MEGA7, com parâmetros padrão.

Fonte: Elaborada pelo autor.

Ensaio preliminares de interação entre as construções P ou HP e G, através termoforese em microescala (MST), mostraram que o domínio GGDEF isolado é capaz de interagir com o domínio PAS (Figura 78). Esse resultado fornece suporte para função estrutural do domínio PAS sobre os estados conformacionais de PA0847. Entretanto, construções isoladas de PAS e GGDEF, quando colocadas juntas no meio reacional, não apresentaram atividade em ensaios com EnzChek (Invitrogen) (dados não mostrados). Isso sugere a necessidade da presença da região conectora entre esses domínios, comumente referenciada como hélice J α .

Figura 72 - Experimento de MST utilizando HP e P marcadas com o fluoróforo e variando a concentração da construção G.

O ensaio foi feito com P ou HP marcadas com nhsNT647 e variando a concentração da construção G não marcada.

Fonte: Elaborada pelo autor.

3.1.8 Caracterização funcional da interação entre PA0847 e HsbD

Uma vez identificada a interação entre PA0847 e HsbD-Nt, em um trabalho anterior do nosso grupo de pesquisa, buscou-se realizar experimentos para confirmar a mesma. Além do mais, buscou-se outros experimentos que pudessem ajudar a compreender o papel da interação no contexto biológico, utilizando construções menores de PA0847.

3.3.8.1 Análises computacionais do segmento N-terminal de HsbD

Num primeiro momento, decidiu-se investigar a relevância da sequência da HsbD-Nt. Foi realizada buscas recursivas com padrões de HMM com a versão *online* do programa

jackhmmer, utilizando parâmetros padrão e o banco de dados Uniprot. Após 7 interações, as buscas convergem para um grupo de 296 sequências das quais 169 (Figura 73), quase a metade, são da ordem *Pseudomonadales*, na qual se insere *P. aeruginosa*.

Figura 73 - Agrupamento de sequências resultantes de acordo com a arquitetura de domínios (A) e padrão do alinhamento das sequências recuperadas (B). A) Agrupamento das sequências em função da arquitetura das proteínas recuperadas das buscas por similaridade pelo jackhmmer, iniciadas com HsbD-Nt. O padrão final é encontrado em proteínas semelhantes a HsbD, com 6 hélices transmembrana, seguidas de domínio GGDEF. De maneira inesperada, aproximadamente 10 % das sequências tem, no lugar do GGDEF, domínios característicos de sistemas sensores HK; B) O padrão de aminoácidos do alinhamento, montado com o programa Skylign. O tamanho do caractere, que representa cada um dos resíduos em cada posição do alinhamento, é igual ao conteúdo de informação calculado para o mesmo. O conteúdo de informação é também chamado entropia relativa ou distância Kullback–Leibler.²⁵² Quando expressa na base $\log_2$, a unidade do conteúdo de informação é dado em bits. Alguns dos resíduos relevantes, segundo o parâmetro de conservação, estão destacados em verde.

Fonte: (A) Adaptada de HMMER...;²¹⁵(B) Adaptada de JODY CLEMENTS.²¹⁷

Todas sequências recuperadas estão localizadas no N-terminal, seguidas de 6 regiões preditas serem hélices transmembrana. Essas regiões transmembrana foram preditas submetendo em batelada (*batch mode*) no programa Phobius as sequências recuperadas das buscas com o jackhmmer. Das proteínas resultantes da busca, a grande maioria, 252 sequências, possuem na região C-terminal um domínio GGDEF. As outras possuem domínios típicos de HK, como DHp (HisKA), CA (HATPase_c) ou REC (Response_reg), como mostrado na Figura

73B. Partindo do alinhamento foi desenhado o padrão de conservação, mostrados na Figura 73A, elaborado pelo servidor *online* Skyline.

A conservação dessa região em diferentes proteínas e com arquitetura de domínios, demonstra a relevância da região. Para determinar quais resíduos do padrão de conservação são chave para a interação, experimentos de mutação sítio dirigida em HsbD de resíduos importantes na conservação (Figura 73) foram iniciados, mas não foram concluídos

3.3.8.2 Confirmação da interação através de ensaios de pull-down

Os experimentos de pull-down foram realizados de acordo com o descrito na seção 2.10. Dos resultados, imobilizando ou a construção HPG-1 (resultados não mostrados) ou a HsbD-Nt fusionadas com SUMO-6xHis (SUMO-Nt) é possível a parceira de interação (Figura 74).

Figura 74 – Pull-down imobilizando HsbD-Nt fusionada a SUMO-6xHis (SUMO-Nt) e recuperando a HPG-1. O gel da esquerda mostra o controle das proteínas usadas, HPG-1 e 6xHis-SUMO-HsbD-Nt (SUMO-Nt). Como descrito (seção 2.10), foram preparadas 5 colunas de afinidade. Em segundo experimento simultâneo, não foi imobilizado a HsbD-Nt (controle sem SUMO-Nt). Em uma terceira resina, não foi aplicada a proteína HPG-1 (SUMO-Nt apenas). Num terceiro experimento, houve a realização da incubação de HPG-1 com resina já contendo a SUMO-Nt imobilizada (HPG-1). No quarto experimento a incubação foi feita com HPG-1 e c-di-GMP (HPG-1 + c-di-GMP). Na última afinidade, além da HPG-1, foi adicionado GTP na incubação e ainda $MgCl_2$, cofator de DGCs. No gel, os controles do não ligado (NL) e da lavagem (LV) quando se utiliza apenas HPG-1 foram utilizados. Além disso, são mostradas as eluições (abreviação E) dos cinco experimentos citados. A recuperação da proteína HPG-1 sem a cauda 6XHis da HPG-1 é explicada com a interação com a HsbD-Nt imobilizada. A maior intensidade da banda da HPG-1 foi provida dos experimentos feitos na presença de c-di-GMP, o que pode ser reflexo da importância da molécula para a interação. Isso mesmo com GTP, uma vez que é sabido que HPG-1 é uma DGC ativa. Abreviações: MM – marcador molecular; NL – não ligado, LV –lavagem; E1 – eluição depois da incubação; E2 – segunda eluição (idêntica a primeira);

Fonte: Elaborada pelo autor.

O resultado reforça a existência entre uma interação entre HPG-1 e HsbD-Nt. A intensidade da banda correspondente a HPG-1 recuperada na eluição é elevada quando se adiciona, junto com HPG-1, c-di-GMP ou GTP com $MgCl_2$. Isso fornece indício da importância

do nucleotídeo para a interação. Foi observado que HPG-1 é instável com c-di-GMP, como visto em análises de SEC-MALS com o nucleotídeo (seção 3.1.4.3). Essa instabilidade é aumentada ainda mais com a adição de SUMO-Nt, de acordo resultados preliminares com ITC, realizados com HPG-1 e c-di-GMP e titulando SUMO-Nt (resultado não mostrado). Nesses experimentos não houve detecção da interação e atribuiu-se a precipitação o resultado negativo.

3.3.8.3 Mapeamento de domínios de PA0847 envolvidos na interação com HsbD-Nt através de ensaios duplo híbrido

Esses experimentos contaram com a colaboração dos mestres Andrea Cardoso e Raphael Meneghello. Os procedimentos experimentais estão detalhados na seção 2.9. Foram realizadas clonagens de construções de PA0847 no vetor pTRG, com mesma abrangência daquelas caracterizadas *in vitro*. Os procedimentos para clonagens seguiram, de acordo com o descrito na seção 2.2.7 e 2.2.8. Se a interação ocorre, a célula cotransformada com as partes que se está testando, uma no vetor pTRG e outra no pBT, sobrevivem e crescem tanto no meio seletivo (com 3-AT) quanto no duplo-seletivo (com 3-AT e estreptomicina).

Nos testes foram reproduzidos os resultados de interação originais entre PA0847-HPG-1, alocado no vetor pTRG, com os N-terminais tanto de PA1851 e HsbD (PA3343), alocados no vetor pBT.¹³ Foi reproduzido também o maior desempenho (maior crescimento) utilizando pBT-HsbD-Nt que com pBT-PA1851-Nt, o que indica uma interação mais forte entre HsbD-Nt com HPG-1 (Figura 75).

Fonte: Elaborada pelo autor.

Devido à proximidade com a membrana da região N-terminal de HsbD, era esperado que o domínio HAMP era o responsável pela interação com HsbD-Nt. Entretanto, para

fragmentos foi observado interação com HsbD-Nt com as construções menores P, PG e HPG-3. Partindo desses resultados pode-se especular que a interação é dependente da presença do domínio PAS, ou mesmo ocorre através desse domínio. E isso condicionado a presença do domínio GGDEF, dado que no experimento com HP não foi observado colônias e que a construção HPG-1 apresentou o crescimento mais evidente dos testes (Figura 75).

Outros experimentos de interação utilizando pETDuet não foi possível observar a banda correspondente a HsbD-Nt, entretanto o resultado negativo foi atribuído a massa pequena da proteína somado ao problema das duas populações de HsbD (dados não mostrados)

3.3.8.4 Mapeamento de domínios de PA0847 envolvidos na interação com HsbD-Nt através de ensaios de MST

Para confrontar os resultados de duplo híbrido e *pull-down*, foram realizados experimentos *in vitro*, através da técnica da termoforese, como descrito na seção 2.8. As construções de PA0847 foram marcadas com os fluoróforos nhsNT647 e mNT647. Marcando a construção a construção HPG-1 com nhsNT647, confirmamos que HsbD-Nt interagem com afinidade em torno de 2 μ M com PA0847 (Figura 76A). O resultado foi similar quando se usou pico 3 ou pico 4 da purificação de HsbD-Nt (Figura 55C). Foi escolhido trabalhar com os experimentos subsequentes com a fração do pico 3 da purificação.

Nos ensaios de *pull-down* verificou-se uma dependência da interação com o c-di-GMP, buscou-se comprovar esse fato por MST. Para evitar que o resultado do ensaio fosse comprometido devido a potencial marcação da arginina do sítio inibitório (RXXD), nos ensaios com c-di-GMP, a construção HPG-1 foi marcada com mNT647. Não foi possível observar diferenças na constante de dissociação entre os ensaios com ou sem o nucleotídeo (Figura 76B). Apesar dos resultados contraditórios entre as duas técnicas, o fato da necessidade da incubação *in vitro* de HPG-1 com o c-di-GMP (que eleva a agregação/precipitação da amostra) dificultou a realização do experimento e talvez tenha comprometido o resultado. Já no ensaio de *pull-down*, a dinâmica é totalmente, pois a interação é feita utilizando proteína imobilizadas.

Focando descobrir qual a dependência da interação observada em função dos domínios da PA0847, foram realizados experimentos de termoforese com construções de PA0847 marcadas com nhsNT-647, usando HsbD-Nt como ligante. Além de HPG-1, uma resposta de interação confiável veio apenas com as construções HPG-2 e HP (Figura 77). Enquanto a afinidade de HsbD-Nt por HPG-2 é similar a calculada para HPG-1, a ausência do domínio GGDEF (construção HP) diminuiu a interação, uma ordem de grandeza, indicando que, apesar

da interação envolver as porções HAMP e PAS de PA0847, o domínio GGDEF, no contexto da conformação da porção citoplasmática da proteína, é importante para interação com HsbD. A não detecção da interação de HsbD-Nt com a construção PG suporta essa hipótese.

Figura 76 - Curvas de MST utilizando HPG-1 marcada e variando a concentração de amostras de HsbD-Nt.

(A) Foram testadas diferentes as frações de HsbD-Nt isolada, com a finalidade de comprovar sua equivalência nos experimentos; (B) Para a presente abordagem experimental não foi possível comprovar a influência do c-di-GMP para interação.

Fonte: Elaborada pelo autor.

Nos ensaios de duplo-híbrido foram observados resultados de interação com as construções P e PG, não observados por MST. Isso pode estar relacionado com as diferenças das técnicas. Enquanto que a marcação com o fluoróforo potencialmente pode influenciar as interações, afetando os valores de K_d obtidos por MST, os experimentos de duplo híbrido são intrinsicamente muito mais complexos e não controlados devido aos componentes do meio de

cultivo e da bactéria utilizadas. Diferenças entre resultados do MST e duplo híbrido podem ser consequência de uma possível sensibilidade de algumas das construções com a modificação covalente provida da marcação com o fluoróforo.

Figura 77 - Curvas de interação entre construções PA0847 marcada com NT-647-NHS e adicionando diferentes concentrações de HsbD-Nt (pico 3 da SEC). Aqui foram plotados $\Delta F_{\text{norm.}}$ pois foi notado que, dependendo da construção, a migração era desfavorecida (HP) ou favorecida (HPG-1 e HPG-2), com a adição do ligante. Experimentos para investigar esse comportamento não foram realizados. Os dados computados para HPG-3, P, PG e G forneceram um sinal de amplitude ($\Delta\Delta F_{\text{norm.}}$) pequenos para validação, o que implicou em dados distorcidos de K_d e de valores estimados de $\Delta\Delta F_{\text{norm.}}$. Por isso esses dados foram descartados.

Fonte: Elaborada pelo autor.

3.3.7.5 Impacto da interação na atividade DGC de PA0847

Para verificar se a interação de HsbD-Nt com HPG-1 impacta na atividade DGC dessa última, ensaios de atividade foram realizados utilizando concentrações crescentes de HsbD-Nt (Figura 78).

Devido a maior estabilidade da construção HsbD-Nt fusionada com SUMO, os ensaios sistemáticos foram utilizados com essas amostras. Destacando que ensaios piloto com a proteína HsbD-Nt isolada apresentaram os mesmos resultados. Mesmo em concentrações mais elevadas de HsbD-Nt, a atividade da HPG-1 não é alterada.

Figura 78 - Teste de atividade de HPG-1 na presença de diferentes concentrações de 6xHis-SUMO-HsbD-Nt (SUMO-Nt). Mesmo com o acréscimo de SUMO-Nt não houve variação da atividade de HPG-1. Isso indica a baixa ou nula interferência da interação na síntese de c-di-GMP por HPG-1. Resultados semelhantes foram obtidos em testes com HsbD-Nt isolado (dados não mostrados).

Fonte: Elaborada pelo autor

3.1.9 Propostas de mecanismos de ativação e funcional do sistema PA0847/HsbD

Baseados nos resultados de oligomerização, atividade e análises computacionais dos domínios de PA0847, bem como o impacto da interação de HsbD-Nt na estrutura e atividade de PA0847, foram levantadas hipóteses sobre a dinâmica regulatória entre PA0847 e HsbD. Ao final, foi gerado um modelo que integraria os resultados obtidos (Figura 82). Embora muitas das interpretações estejam de acordo com os resultados experimentais, o modelo apresentado envolve especulações, que necessitariam de mais evidências experimentais para serem melhor suportadas.

Para facilitar as visualizações, foram elaboradas esquematizações para o dímero GGDEF ativo, baseado na estrutura de DgcZ, e inativo (modo inibido 1), baseado na estrutura de PleD (Figura 79). Essas são as únicas formas encontradas em estruturas cristalográficas de dímeros de GGDEFs já reportadas, nas quais o controle da dimerização é atribuído aos domínios regulatórios ou pelo c-di-GMP (Figura 79).

O domínio GGDEF isolado de PA0847 foi responsável em equilíbrio monômero-dímero na forma apo. E isso é um fato inédito, dado que esse domínio isolado apenas foi reportado sem c-di-GMP, até o momento, como monômeros. É possível que interface de dimerização entre GGDEFs de PA0847 ocorra numa conformação parecida com a forma inibida 1. Isso porque formas oligoméricas se mostraram dependentes do sítio inibitório RXXD (sítio I). Não deve ser exatamente a mesma interface porque as construções não copurificaram com c-di-GMP, como HPG-1, HPG-2 e G. Construções como HPG-1, PG e G apareceram instáveis com o nucleotídeo, o que pode indicar que a forma inibida 1, induzida pelo c-di-GMP, pode ter sido responsável pela instabilidade das construções (Figura 80).

A copurificação de c-di-GMP foi observada depende da inclusão na construção do domínio PAS. Isso leva a acreditar que exista um mecanismo de ligação intracadeia, como a observada para a modo de inibição 2 (Figura 14) ou um modo de inibição 1 dependente da conformação dos domínios HAMP e PAS. A tendência para copurificação do c-di-GMP para as construções, assim como a dificuldade em remover o nucleotídeo, foi, da maior para menor: PG > HPG-3 > HPG-2. Isso permite acrescentar que o efeito que o domínio HAMP imprime no domínio PAS foi gradualmente negativo na ligação ao c-di-GMP.

Figura 79 - Representação das estruturas do domínio GGDEF na conformação ativa e inibida (acima) e das representações simplificadas (abaixo).

Cada unidade do dímero está representada com uma tonalidade de verde diferente. O prolongamento destacado na esquematização do domínio GGDEF é composto pelas hélices $\alpha 2$ e $\alpha 2'$ e $\beta 3'$ e $\beta 3''$. Essa estrutura fecha lateralmente o sítio ativo, na conformação ativa, como mostrado. Para haver uma mudança da conformação ativa para a inibida, sem trocar a região N-terminal de lugar, é preciso rotacionar o domínio GGDEF, afastando os sítios. Na estrutura inibida não foi representado o c-di-GMP por questões de clareza. Em destaque, na esquematização da forma inibida, é mostrado o sítio de ligação do dímero de c-di-GMP.

Fonte: Elaborada pelo autor.

As variações das formas globulares e alongadas observadas nos experimentos de SEC-MALS e com nucleotídeo podem ser decorrentes de mudanças conformacionais promovidas pelo domínio HAMP (HPG-1 mais globular e HPG-2 e HPG-3 mais alongadas) e pela ligação ao c-di-GMP (PG mais alongada quando ligada e mais globular quando na forma apo). Entretanto, para explicar como esses fenômenos ocorrem e quais são suas implicações, mais experimentos seriam necessários.

Figura 80 - Formas diméricas possíveis das construções G, PG e HPG-2 ou HPG-3.

Os dímeros das construções G, PG, HPG-2 e HPG-3 representados, segundo o modelo proposto, depende do motivo RXXD. O domínio PAS deve conferir a propriedade de ligar o c-di-GMP (seta vermelha), ligando diretamente o nucleotídeo, ou indiretamente, alterando a forma que o domínio GGDEF liga c-di-GMP. Na esquematização, as unidades do domínio PAS estão representados em tonalidades de azul. Os GGDEF são apresentados em tonalidades em verde. As hélices do HAMP estão representadas por bastões.

Fonte: Elaborada pelo autor.

Para explicar a atividade das construções PG, HPG-2 e HPG-3, em algum momento essas construções devem formar um dímero ativo, provavelmente similar a conformação observada para GGDEFs ativos clássicos. É coerente propor para isso que ocorra um encontro de dímeros dessas construções (Figura 80), formando um tetrâmero transiente ativo, ao invés de uma mudança conformacional de um dímero inativo para um ativo (Figura 79). Entretanto, é necessário também assumir que o tetrâmero de transição rápida (HPG-2, HPG-3 e PG) e o tetrâmero em equilíbrio (HPG-1) devem ter dinâmicas diferentes, dado que o aumento da concentração de NaCl apenas alterou a atividade de HPG-1 (equilíbrio), mas não na atividade de HPG-2, HPG-3 e PG.

A formação de um tetrâmero transiente exige uma mudança conformacional não tão drástica das unidades do dímero inativo uma rotação de $\approx 45^\circ$ em um eixo de cada uma das unidades do dímero não ativo (Figura 81A). Por outro lado, para obter um dímero inibido partindo de um dímero ativo, e vice-versa, é necessária uma rotação de $\approx 180^\circ$ (observando apenas um eixo) (Figura 79). Além disso, a forma tetramérica foi apontada como a responsável pela maior atividade da construção HPG-1. Essa alternância de formas apenas foi possível com adição do domínio PAS, uma vez que a construção G não apresentou atividade. Isso reforça o papel estrutural do domínio PAS e fornece um exemplo da importância dos domínios regulatórios para GGDEF.

O HAMP pode estar contribuindo para estabilização do tetrâmero ativo, uma vez que aparecem tetrâmeros e com elevada atividade para construção HPG-1 (Figura 81B). Nesse

através da rotação dos GGDEF no próprio eixo (Figura 82B). Portanto, no tetrâmero, duas das interfaces são responsáveis pela atividade, pela formação do sítio A (A), enquanto a outra é responsável pela interface do sítio I (I).

Figura 82 - Mecanismo para PA0847 para formação de tetrâmeros (A) e sua alternância entre estados ativos e inativos (B).

A) Ilustração da vista lateral e superior (não mostrando GGDEF) do tetrâmero da porção citoplasmática de PA0847 na forma ativa. Os domínios GGDEF na figura são destacados em verde, enquanto os domínios PAS estão em azul. As hélices do domínio HAMP estão como bastões rosa e azul. As cadeias em cinza correspondem a unidades diméricas simétricas. As hélices AS1 e AS2 do domínio HAMP foi simplificada para numeração 1 e 2; B) Representação da alternância de conformações ativas e inativas no tetrâmero. O modelo de bolas abaixo representa um modelo geral para as duas formas. No tetrâmero, as faces de contato marcadas como I são responsáveis pela estabilização dependente do sítio I. Outros contatos, marcados com a letra A, são onde pode ocorrer a forma ativa. Na esquematização, as unidades do domínio PAS estão representados em tonalidades de azul. Os GGDEF são apresentados em tonalidades em verde. As hélices do HAMP estão representadas por bastões.

Fonte: Elaborada pelo autor.

Os resultados de MST, duplo híbrido e *pull-down* concordam que ocorre interação entre HPG-1 e HsbD-Nt. Resultados do MST apontam a interação com HsbD-Nt com o domínio HAMP, interação intensificada com a presença do domínio GGDEF. Por outro lado, o duplo híbrido suporta a interação entre HsbD-Nt com o domínio PAS, também intensificada com a presença na construção do domínio GGDEF. Uma possível interpretação, portanto, foi que a interação ocorre com os domínios PAS e HAMP, condicionada à presença do domínio GGDEF e ainda a algum fator adicional, provindo de diferenças entre esses experimentos. Então a interação deve ocorrer com todos os domínios, mas como não foi observada a interação em nenhum experimento com o GGDEF isolado, resta a dúvida se ele participa da interação diretamente ou alterando os domínios PAS e HAMP.

Como HsbD não possui efeito sobre a atividade de PA0847, a interação pode ser usada para colocalizar PA0847 e HsbD. Essa colocalização pode estar envolvida em um processo de

direcionamento específico de c-di-GMP, como visto em outros exemplos.⁹⁻¹² Nesse contexto, outra hipótese seria a presença de algum sinal para que ocorra alguma alteração na atividade de PA0847 ou de HsbD ou mesmo para desfazer um complexo entre essas proteínas que pode estar voltado para um controle específico de c-di-GMP (Figura 83).

Figura 83 - Mecanismo ilustrativo da sinalização para PA0847 e HsbD.

A interação pode ser responsável por colocalizar na célula as proteínas. Uma modificação estrutural, provinda do domínio periplasmático de PA0847, ou das hélices transmembrana de HsbD pode interferir na interação, desfazendo o complexo, por exemplo, ou pode alterar a atividade das proteínas.

Fonte: Elaborada pelo autor

3.2 Análises de coevolução indicam novos resíduos chave para GGDEF-EALs

Devido à importância das proteínas com os domínios GGDEF-EAL foram realizadas análises de sequências com essa arquitetura, a fim de obter pistas sobre como esses sistemas funcionam e como a presença simultânea desses domínios poderia ser vantajosa nas vias do c-di-GMP.

3.2.1 Análises de correlação pela métrica de perturbação proposta por Bleicher e colaboradores

Para a investigação de resíduos chave de proteínas com domínios GGDEF e EAL em série, foram feitas análises em larga escala de sequências com essa configuração. Inicialmente, procedeu-se com a busca por sequências nos bancos *online*. É apresentado no banco de Galperin e colaboradores^{91,200} um total 5153 sequências com tanto GGDEF e EAL. Dado o procedimento de filtragem, como enumerado na seção 2.12, foram recuperadas 4399 sequências (que não possuíam links do Uniprot obsoletos e desvios da sequência e no alinhamento). Já o banco Uniprot¹⁷⁴ apresenta 80707 sequências com GGDEF e EAL, das quais 73239 passaram pelos filtros de seleção, (seção 2.12). Como se tratam de diversas sequências não revisadas, sobre essas foi aplicado um filtro por identidade e cobertura, utilizando o programa PFstats.²⁰¹⁻²⁰² Foram removidas as sequências, com identidade maior que 80 %, menor que 15 % e com cobertura menor que 80 % em relação a sequência da STM3615 (código Uniprot Q8ZLB9_SALTY). Ao final das filtrações, foram obtidas 27004 sequências do Uniprot.

Com os alinhamentos finais obtidos com os bancos de Galperin e do Uniprot foram feitas classificações, em função da conservação dos domínios GGDEF e EAL e também em relação ao motivo RXXD, de acordo com o definido na seção 2.12, Tabela 11. Esses dados de conservação foram analisados graficamente (Figura 84) e estão resumidos na Tabela 22. O grupo de sequências, tanto o GGDEF quanto o EAL, contendo resíduos para catálise foi nomeado G&E. As sequências com o GGDEF conservado e EAL predito inativo foram alocados no grupo G. As sequências nas quais ocorreu o contrário, apenas o EAL predito ativo, foi classificado no grupo E. O grupo de sequências sem resíduos importantes tanto para catálise DGC e PDE foram agrupados como --. Os grupos obtidos com o Uniprot e Galperin ficaram próximos em proporção de sequências, o que suporta uma coerência entre esses dados. As linhas de código e resultados detalhados dessa análise se encontram disponíveis *online*.²⁵³

Foi observado que as sequências com EAL predito inativo seguido do GGDEF predito ativo são uma minoria, compondo aproximadamente 3 % do total das sequências recuperadas. A identificação de uma pequena porção de sequências com GGDEF e EAL degenerado (grupo --), aproximadamente 7 %, pode ser consequência natural da dificuldade em identificar esse tipo de sequências, frente a degeneração dos resíduos catalíticos que caracterizam os domínios. Muitas destas sequências foram eliminadas com o filtro de espaçamento no alinhamento, que não podia ser maior que 200. Para as sequências derivadas do filtro de Galperin e colaboradores^{91,200} por exemplo, 196 sequências em relação ao GGDEF e 28 sequências em relação ao EAL não respeitam o critério. Esse número se equipara ao número de sequências do grupo --. Esses fatos indicam a necessidade de uma abordagem particular para esse tipo de sequências.

Tabela 22 - Distribuição das sequências recuperadas nos bancos de dados de acordo com a conservação dos domínios GGDEF e EAL e da conservação do motivo RXXD

Grupos	Galperin			Uniprot		
	Total	RXXD conservado	RXXD não conservado	Total	RXXD conservado	RXXD não conservado
Total	4399 (100 %)	1726 (39,2 %)	2673 (60,2 %)	27004 (100 %)	10481 (38,8 %)	16523 (61,2 %)
G&E	2622 (59,6 %)	1371 (52,3 %)	1251 (47,7 %)	15700 (58,2)	8120 (51,7 %)	7580 (48,3 %)
G	148 (3,4 %)	115 (77,7 %)	33 (22,3 %)	974 (3,6 %)	668 (68,6 %)	306 (31,4 %)
E	1318 (30,0 %)	212 (16,1 %)	1106 (83,9 %)	8499 (31,5 %)	1488 (17,5 %)	7011 (82,5 %)
--	311 (7,1 %)	28 (9,0 %)	283 (91,0 %)	1831 (6,8 %)	205 (11,2 %)	1626 (88,8 %)

*G&E: grupo com GGDEF e EAL conservado; G: grupo com apenas o GGDEF conservado; E: grupo com apenas o EAL conservado; --: grupo que possui nem o GGDEF e nem o EAL conservado.

Fonte: Elaborada pelo autor.

A queda na incidência do motivo RXXD no grupo G&E (≈ 50 %) em relação ao grupo G (≈ 75 %) pode ser indício da presença de outros mecanismos para regulação da atividade DGC, além daquele descrito para o RXXD. Como proposto em um estudo anterior, uma hélice que conecta os domínios GGDEF e EAL teria a função alterar os estados dessas proteínas com pelo menos um dos domínios ativos.¹⁶

Devido ao banco de dados de Galperin ser melhor catalogado e por restringir a sequências revisadas, o alinhamento deste banco foi escolhido para análises de coevolução de resíduos. A análise foi feita com o objetivo de encontrar comunidades de resíduos que fossem próprias de grupos de sequências:

- que possuíssem tanto o GGDEF quanto o EAL predito ativo (grupo G&E);
- com apenas o domínio GGDEF conservado (grupo G);
- com apenas o domínio EAL conservado (grupo E);
- com nenhum dos domínios conservados (grupo --).

Outra comunidade de resíduos que se buscou foi aquela que poderia explicar se uma sequência é intrinsecamente dimérica ou não. A caracterização de uma comunidade através da relação funcional dos resíduos que compõem a comunidade foi abordada no tratamento das proteínas da família de superóxido dismutases (SOD), na qual as comunidades se correlacionaram com a oligomerização e especificidade por metais ferro ou manganês dessas proteínas.^{201,254}

Figura 84 - Distribuição das sequências recuperadas divididos em relação a conservação do sítio inibitório RXXD.

Fonte: Elaborada pelo autor.

Nos cálculos de correlação, foram utilizados como corte as frequências condicionais maiores 30 % e anticorrelações menores que 70 %. Isso para tentar aumentar as chances de recuperar resíduos que sejam importantes para subgrupos de sequências, mas com representatividade pequena. O corte da pontuação logP escolhido foi de 70, de maneira que fossem recuperadas aproximadamente as 50 correlações mais expressivas, número arbitrário.

Foram encontradas 12 comunidades (Tabela 23, Figura 85), sendo que cada comunidade teve de 16 a 2 resíduos. Os resíduos nas colunas da Tabela 23 estão apresentados, da esquerda para direita, de acordo com a importância do resíduo correlacionado para a comunidade. As tabelas de variação de pontuação e de frequência dos resíduos recuperados nas comunidades estão no apêndice D.

Em uma análise adicional, as frequências de qualquer resíduo, de cada uma das posições identificadas nas comunidades encontradas, foram recuperadas. Isso foi feito usando linhas de código em Python. O alinhamento foi dividido em sub-alinhamentos, de acordo com os grupos de conservação dos domínios GGDEF e EAL (G&E, G, E e --). Desses sub-alinhamentos, também foram calculadas as frequências de qualquer resíduo, nas mesmas posições recuperadas nas comunidades. Esses dados foram usados para elaborar a Tabela 23. Se a frequência de um resíduo no sub-alinhamento era maior, em pelo menos 1 %, em relação ao alinhamento total,

esse resíduo da rede era classificado com pertencente ao grupo de conservação. Essa divisão é mostrada na rede apresentada na Figura 85. A Tabela 23 mostra também se outro resíduo, além daquele encontrado por correlação, possui uma variação de frequência maior no processo de sub-alinhamento. A classificação em grupos de conservação foi uma tentativa relacionar as comunidades com possíveis funções de acordo com a conservação.

As baixas flutuações das frequências dos resíduos no alinhamento total para cada os grupos de conservação pode ser reflexo da heterogeneidade das proteínas utilizadas para análise. Ou seja, apenas algumas das proteínas do grupo utilizam os resíduos de alguma forma. Usando o PFstats com alfabetos reduzidos²⁵⁵ (considerando resíduos equivalentes se compartilham propriedades físico-químicas), não foi obtido grandes variações nos resultados de correlação.

Para exemplificar a análise, tomemos como exemplo os resíduos K659(S) e D682(E), da comunidade 6. O PFstats identificou que um resíduo K na posição do alinhamento 1170 possui uma correlação com um D na posição 1243. A pontuação da correlação, $\log(P)$, para esse par foi 141 (Tabela 32). Essas posições correlacionadas correspondem às posições 659 e 682, respectivamente, da sequência de TBD1265. Entretanto, na referência são encontrados um resíduo S na posição 659 e um resíduo E na posição 682. Pelos resultados do PFstats, foi possível observar que as sequências com K659 conservado, 68,6 % apresentam também um D682 (ver as tabelas do apêndice D para resultados completos). No alinhamento geral, 37,3 % das sequências possuem D682 (um D na posição 1243 do alinhamento). No sub-alinhamento de sequências com GGDEF e EAL conservado, a frequência de D682 é de 43,8 %, o que representa um aumento na incidência de 6,5 %. Para outros grupos houve uma queda da frequência de D682, por isso, esse resíduo foi classificado como pertencente ao grupo G&E, como mostra a rede da Figura 85. Apesar disso, o resíduo N, na mesma posição de D682, teve uma diferença de frequências entre o alinhamento total e o grupo -- em 30,6 % (de 11,5 % para 42,1 %). Essa variação é maior, em valor absoluto, que o observado para o resíduo D (de 37,3 % para 7,4%, uma diferença de -29,9 %).

Tabela 23 - Frequência dos resíduos (%) com correlação e variação da frequência em relação aos subgrupos de acordo com a conservação dos domínios GGDEF (abreviação G) e EAL (abreviação E).

	Comunidade 1									
Resíduo	K332(K)	Y439(Y)	H340(H)	D435(D)	D336(D)	E371(E)	D327(D)	G368(G)	G369(G)	S414(S)
%	74,9	61,3	73,4	72,9	77,3	80,4	79,8	73,1	80,0	79,2
G&E	20,5	24,8	22,8	21,2	19,2	19,6	17,0	19,6	20,1	16,6
E	-28,3	-40,3	-34,1	-37,7	-26,4	-29,3	-21,9	-28,7	-31,2	-23,0
G	22,4	32,7	22,6	25,1	20,7	19,6	20,2	18,1	20,1	18,1
--	-64,0	-54,2	-58,3	-31,1	-60,2	-50,5	-59,9	-52,6	-46,2	-51,3
Resíduo	M438(V) F325(L)* D329(D) L328(I)				L438(V)*		A373(V)*			
%	50,8	35,6	80,3	67,2	35,1		34,8			
G&E	14,0	13,7	19,7	15,6	-6,2/[M]:14,0		-6,42/[V]:6,8			
E	-20,7	-22,5	-30,7	-26,1	10,9/[M]:-20,7		7,1/[V]:-9,7			
G	15,4	17,8	19,7	20,6	-11,4/[M]:15,4		-3,7/[V]:8,5			
--	-37,7	-28,8	-45,5	-30,9	11,9/[M]:-37,7		26,0			
	Comunidade 2									
Resíduo	R359(R)	D362(D)	T363(L)	R306(Q)	L302(Y)	R316(L)	A310(L)	P298(H)		
%	42,5	52,2	31,6	39,2	31,8	30,6	36,9	57,9		
G&E	13,1	13,3	9,0	9,4	8,2	9,0	8,5	5,8		
E	-23,3	-21,9	-14,5	-17,2	-12,2	-14,2/[-]:15,1		-15,4	-3,6/[A]:-4,2	
G	37,9	31,5	13,7	16,9	7,4/[M]:8,7 10,0/[-]:-10,1		21,9	1,5/[A]:10,4		
--	-29,6	-33,9	-20,6	-14,1/[Q]:18,5	-20,9	-20,3	-16,6	-34,2		
	Comunidade 3									
Resíduo	H530(D)	P531(P)	G534(G)	W528(W)	C719(C)	G704(G)	W569(W)	Q565(Q)		
%	55,8	66,4	73,4	80,8	71,8	88,2	66,8	43,9		
G&E	6,5	6,0	5,8	5,4	6,0	3,5	8,0	6,1		
E	-4,7	-2,2	-1,3	1,1	1,2	0,2	-4,9	-5,3		
G	3,7	-2,2	-3,2	5,7	-15,0	-7,8	-16,8	-4,0		
--	-37,1	-40,4	-42,2	-52,8	-48,6	-26,8	-39,1	-27,2		
	Comunidade 4					Comunidade 5				
Resíduo	G555(S)	Q589(Q)	C563(C)	I542(I)	I551(I)	P421(A)	D423(E)	G424(G)		
%	60,8	57,9	53,3	72,3	69,7	55,6	39,2	42,9		
G&E	9,8	9,4	8,1	6,2	6,3	10,4	8,0	4,4		
E	-6,8/[T]:7,6		-6,5/[D]:9,0		-5,8	3,9	2,2/[Y]:3,2	-10,6	-8,6	-8,2
G	-12,2		-16,0		-8,0	-50,0	-29,8	-8,9	-8,7	-0,4/[A]:-7,0
--	-48,0		-43,7		-39,4	-44,7	-48,1	-37,9/38,7[-]	-26,9	-2,4/[A]:-6,1
	Comunidade 6		Comunidade 7		Comunidade 8		Comunidade 9			
Resíduo	K659(S)	D682(E)	D370(E)	K442(K)	N335(N)	G343(G)	G445(G)	N447(N)		
%	36,7	37,3	85,9	85,6	90,0	87,4	74,7	43,4		
G&E	4,1	6,5	12,6	13,9	9,9	11,5	13,2	10,4/[-]:-15,2		
E	-3,6	-5,0	-12,1	-14,0	-10,3	-11,6	-22,5	-17,2/[-]:24,5		
G	-20,5		-8,2		13,4	13,8	10,0	12,6	17,8	-1,5/[-]:-9,1
--	-9,7	-29,9/[N]:30,6		-61,8	-64,3	-44,7	-53,6	-23,9	-14,2/[-]:28,2	
	Comunidade 10			Comunidade 11			Comunidade 12			
Resíduo	L506(L)		W557(W)	L585(I)*		F595(L)*	I511(I)*	E614(E)*		
%	62,3		62,0	43,2		40,7	30,8	37,2		
G&E	2,7		6,8	-0,8/[V]:2,2		-1,4/[L]:2,9	-0,6/[Q]:3,5		5,9	
E	-5,2		-5,0	-0,9/[I]:1,7		-2,0	1,6/[K]:5,5		-3,6	
G	1,9/[P]:-5,21		-6,6	4,7/[V]:-5,0		-5,6/[L]:-8,5	-0,4/[V]:17,5		-8,9	
--	-1,5/[F]:5,7		-32,4	8,2/[V]:-12,6		23,3	-1,5/[V]:26,4		-30,5	

Resíduos X#(#): resíduo com correlação, posição equivalente nas referências (PleD e TBD1265) e, entre parênteses, o resíduo encontrado nas referências; % : frequência no alinhamento total; Colchetes: resíduo com variação maior que aquela com correlação, na mesma posição; G&E, G, E e -- : referente aos subgrupos de sequências com GGDEF e EAL conservado (G&E), apenas GGDEF (G) ou EAL (E) conservado ou nenhum deles (--); verde: resíduos do GGDEF; laranja: resíduos do EALs; *: resíduos que apresentam anti-correlação.
Fonte: Elaborada pelo autor.

Figura 85 - Rede de correlação e divisão das comunidades de acordo com os grupos de conservação dos domínios GGDEF e EAL.

Os resíduos dentro de círculos verdes pertencem ao domínio GGDEF e em laranja ao EAL. A posição são referentes a TBD1265 (EAL) e PleD (GGDEF). Os resíduos apresentados em parênteses representam aqueles encontrados nas sequências de referência. Os traços são diretamente proporcionais a pontuação para correlação entre os resíduos. A cor azul apresenta uma correlação positiva, ou seja, caso um dos resíduos está presente, as chances de observar o outro é elevada. Os traços em vermelho indicam uma correlação negativa, ou seja, dado a presença de um resíduo, a frequência do outro diminui. As regiões cinza abrangem resíduos da mesma comunidade. Elas etiquetadas de C1 até C12, numeração que corresponde a cada comunidade. As regiões delimitadas por traços abrangem os resíduos que apresentam a mesma tendência pelos grupos de conservação dos domínios GGDEF e EAL (G&E, G, E e --). As redes mostradas foram montadas através de biblioteca do Python, a partir de dados das comunidades resultantes do programa PFstats.

Fonte: Elaborada pelo autor.

Os resíduos de cada uma das comunidades foram identificados nas estruturas das sequências de referência de PleD (PDB: 2v0n) e TBD1265 (PDB: 3n3t). Um modelo dimérico ativo da PleD foi elaborado, fazendo um alinhamento de estruturas pelo programa Pymol²¹⁹, o

GGDEF da cadeia A de PleD com o dímero ativo de GGDEFs encontrado em DgcZ (PDB: 3tvk). A partir das estruturas, hipóteses foram propostas função para cada um dos resíduos de cada uma das comunidades. Um resumo das análises estruturais é apresentado na Figura 86.

Inicialmente, para comunidade 1, os resíduos em grande parte, estão associados com a formação de c-di-GMP ou mesmo para coordenação em receptores de GTP, com é o caso de FimX e CC3396.¹⁰⁶⁻¹⁰⁷ A frequência de resíduos da comunidade aumenta quando o alinhamento é restrito para grupos com GGDEF conservado, que são os grupos G&E e G (Tabela 23). Exemplos claros da relação da comunidade 1 são os resíduos do motivo GGDEF G368, G369 e E371, do resíduo proposto para estabilizar o estado de transição K332, o resíduo que coordena o metal D327 e resíduos do loop P, D329 e L328, conhecido por acomodar fosfatos do substrato. Apesar de distante estruturalmente, S414 se encontra bem próximo ao loop P, talvez participando da função desse loop. O par mais relevante da comunidade 1 são K332 e Y439.

O resíduo Y439 apresenta correlação com 12 dos 16 resíduos da comunidade 1. É necessário observar que os resíduos do loop P, D329 e L328, não possuem correlação entre si, mas sim com Y439. Em um dímero ativo, Y439 se encontra próximo ao mesmo resíduo da unidade vizinha (Figura 86-C1), mas distante do sítio ativo, na hélice $\alpha 4$. Algo semelhante ocorre com H340, porém essa se encontra em $\alpha 2$, na região de especificidade do grupo guanil, próxima ao sítio catalítico e distante de Y439. Os pares de resíduos anti-correlacionados F325(L) e L438(V) ou A373(V) se ligam à rede com Y439 através da correlação com M438(V). As posições 438, 325 e 373 se encontram em $\alpha 4$, $\beta 1$ e $\beta 3$, sendo que $\beta 3$ se localiza o final do motivo GGDEF. Isso sugere uma conexão entre Y439 com o sítio ativo. Já K332, um resíduo carregado positivamente, possui correlação com resíduos de carga negativa D327, D435, da mesma hélice de Y439, e D336, resíduo que divide a hélice $\alpha 1$ com a própria K332.

Devido a relação dos resíduos com K332, pode-se especular que, assim como K332, alguns dos resíduos estejam diretamente envolvidos com passos da catálise. Outra hipótese é a de que esses resíduos estariam envolvidos na estabilização do dímero ativo, dados os possíveis pares hidrofóbicos Y439-Y439' e H340-H340' e os carregados R332-D327' e R332-D435'. Pode-se sugerir ainda uma terceira hipótese, de que um mecanismo de regulação da atividade DGC ocorra através do resíduo Y439, já que ele é central na rede de se encontra exposta ao solvente.

Figura 86 - Análises estrutural das comunidades obtidas do programa PFstats.

Disposição estrutural dos resíduos correlacionados (em preto) distribuídos por comunidades. Em verde está representada a estrutura do domínio GGDEF (referência PleD, PDB: 2v0n) e em laranja esta representada o EAL (referência TBD1265, PDB: 3tvk). Em destaque rosa estão as cadeias laterais dos resíduos correlacionados.

Fonte: Elaborada pelo autor.

Para comunidade 2 pode-se observar a correlação clássica entre os R362 e D359, do sítio inibitório RXXD. Assim como na comunidade 1, os resíduos da comunidade 2 tiveram presença elevada nos grupos G e G&E. Um resíduo central na rede de correlação, T363, parece conectar os resíduos do sítio I com todos os demais resíduos da rede, todos pertencentes a hélice $\alpha 0$, que são P298, L302, R306, A310 e R316. Na hélice $\alpha 0$ se localiza o sítio secundário de inibição 1, que em PleD corresponde a R313. Nesse modo de inibição, ocorre a dimerização entre GGDEFs promovido por um c-di-GMP dimérico. Excluindo o sítio inibitório, nenhum dos demais resíduos da comunidade 2 são encontradas na referência PleD. Considerando que PleD apresenta a regulação os dois modos de inibição regulados por c-di-GMP, pode-se especular que a rede encontrada é referente a um outro tipo de regulação, governado pela hélice $\alpha 0$.

A comunidade 3, por sua vez, abrange resíduos do domínio EAL, com frequência maior, no geral, em sequências com os dois domínios conservados (grupo G&E). Os resíduos mais relevantes da comunidade se encontram colocalizados entre $\beta 2$ e $\beta 2'$, que são W528, G534, D530, P531. Os resíduos W569 e Q565, com menor relevância, se encontram no final da hélice $\alpha 2'$. Todos esses resíduos da comunidade 2 se encontram voltados para mesma região do barril do EAL, exposta ao solvente em um dímero clássico de EALs e próximas as hélices $\alpha 0$ e $\alpha 8$. Por outro lado, C719 e G704 se encontram próximas a interface de dimerização. G704 e C719 se encontram próximos na sequência e na estrutura, respectivamente, do resíduo catalítico E703 e do resíduo de especificidade pelo grupo guanil E706. Pode-se levantar a hipótese de que os resíduos que compartilham a mesma face no EAL sejam responsáveis por captarem e transmitirem uma mudança conformacional para os outros dois resíduos próximos a alguns resíduos importantes para catálise. Essa mudança conformacional pode ser em resposta a uma mudança provinda da hélice $\alpha 0$ ou por um possível parceiro de interação.

Os resíduos da comunidade 4 estão, na maioria, relacionados com o grupo de conservação G&E e estão próximos com a conexão entre $\alpha 2$ e $\alpha 2'$. A correlação de maior peso da comunidade é entre G555, pertencente a $\alpha 2'$, e Q589, localizada em $\alpha 3$. Esses resíduos parecem acoplar, portanto, essas hélices vizinhas. I542 e I551, apesar de não serem tão importantes para a comunidade, são conhecidos por compor um bolsão que acomoda um dos grupos guanil do c-di-GMP. Portanto a comunidade deve estar relacionada a ligação ao c-di-GMP. Além disso, a comunidade pode estar envolvida com uma resposta conformacional em função do nucleotídeo, o que justificaria, por exemplo, a presença de C563, que também esta em $\alpha 2'$, mas esta distante dos demais resíduos da comunidade.

A comunidade 5 envolve resíduos P421, D423 e G424, que possuem uma frequência enriquecida em proteínas do grupo de conservação G&E. Todos esses resíduos se encontram entre $\beta 4$ e $\alpha 4$ na estrutura do domínio GGDEF. Esses resíduos se encontram expostos ao solvente e isso levanta a hipótese de que possam ser utilizados para interação com o domínio EAL em determinada conformação ou para interação com um parceiro desconhecido. Um exemplo parecido e descrito na literatura é da interação da $\alpha 4$ do domínio GGDEF de GcbC e a hélice $\alpha 2$ do domínio EAL de LapD.¹² Outro exemplo de interação, mas que envolve $\alpha 3$ do GGDEF com $\alpha 2$ é MorA. Nessa proteína, o resíduo K1076 da hélice $\alpha 3$ do domínio GGDEF com o resíduo E1213 localizado em $\alpha 2$ do EAL da mesma cadeia polipeptídica, isso apenas na forma apo (PDB: 4rnf).¹¹⁵ Esses resíduos correspondem a R386 de PleD e E546 de TBD1265.

A comunidade 6 ilustra uma interação iônica entre resíduos na interface de dimerização entre EALs, que são K659 da hélice $\alpha 5$ e D682 da hélice $\alpha 6'$. Esses dois resíduos são mais importantes para o grupo G&E, segundo a variação de frequências. A comunidade 7 apresentou o resíduo D370, resíduo catalítico do GGDEF, e o resíduo K442, que ajuda a coordenar os fosfatos β e γ do GTP.⁹⁴ Isso pode explicar a razão desses resíduos serem mais pronunciados em grupos com GGDEF conservado G e G&E. De maneira semelhante, a comunidade 9, que engloba N447 e G445, está próxima a resíduos K442 e R446 que coordenam os fosfatos do substrato.

A comunidade 8 engloba resíduos das hélices $\alpha 1$ e $\alpha 2$ do GGDEF, que são N335, conhecido por conferir especificidade pelo grupo guanil no domínio GGDEF, além do resíduo G343, que também deve participar da interação com o c-di-GMP. Os resíduos devem possuir uma importância associada com a atividade no geral por terem afinidade pelos grupos G e G&E. Já a comunidade 10 faz referência aos resíduos do EAL L506, da fita $\beta 1$, e W557, da hélice $\alpha 2'$, resíduos com as cadeias laterais em contato, o que pode indicar que possuem a função de manter a estrutura. Ambos os resíduos são importantes para o grupo G&E entretanto, apenas L506 é importante ainda para o grupo G. Porém, as variações de frequências associados aos grupos dessa comunidade são baixas, o que leva a questionar a classificação de grupos de conservação para esses resíduos.

As comunidades 11 e 12 correlacionam resíduos em contatos, entretanto essas comunidades foram montadas baseadas em anti-correlação. Na comunidade 11, L585 e F595 aparentam manter a estrutura de $\alpha 3$ e $\beta 3$, esta última abriga N584, um dos resíduos da coordenação do metal divalente Me1. A comunidade 12 possui os resíduos I511 e E614, das fitas $\beta 4$ e $\beta 6$ e ambos na região N terminal do barril. A fita $\beta 4$ reside também E616, um resíduo

que coordena Me1, enquanto $\beta 1$ é a fita invertida particular do EAL e alongada, conectando a região N-terminal do EAL com o núcleo do domínio. A comunidade 12 pode ajudar, portanto, a entender o porquê da fita $\beta 1$ ser invertida e se ela de fato exerce algum papel na regulação do EAL.

3.2.2 Análises da interface dimérica de EALs

Uma das questões do trabalho era buscar entender a vantagem da dimerização constitutiva reportada para proteína STM3615. Outra pergunta é como essa proteína é regulada, uma vez que em outros exemplos a dimerização é promovida apenas pelo c-di-GMP, por mecanismos que envolvem mudanças conformacionais no motivo regulatório DDFG(T/A)GYSS e o resíduo E619 (numeração de TBD1265) do loop entre $\beta 4$ e $\alpha 4$ do EAL.⁹⁰

Não foi possível identificar um grupo de resíduos correlacionados que seja exclusivo do grupo de proteínas GGDEF e EAL que possuem apenas o EAL conservado, grupo poderia trazer pistas de como seria a regulação de STM3615. Portanto realizou-se outra abordagem, iniciada com a análise do alinhamento de EALs já descritos para investigação das peculiaridades do EAL de STM3615 (alinhamento completo no apêndice C, Figura 90).

Foi constatado que STM3615 apresenta todos os resíduos catalíticos do EAL conservados, mas com modificações importantes na interface de dimerização, com uma deleção em $\alpha 5$, uma inserção em $\alpha 6'$ e a ausência dos resíduos finais do motivo regulatório YSS. A região do alinhamento contendo essas regiões estão destacadas na Figura 87). Essas particularidades também são observadas em outras proteínas de microrganismos distintos, como SFV_3559, de *Shigella flexneri*, YhjK ou PdeK, de *E. coli* e HmsP, de *Y. pestis*. Essas proteínas possuem o domínio EAL com atividade PDE constatada pela literatura.^{82,256} SFV_3559, PdeK e HsmP possuem a mesma arquitetura de domínios da STM3615, com um GAPES3 periplasmático e os domínios HAMP, GGDEF degenerado e EAL conservado no citoplasmáticos (de acordo com o predito pelo Pfam, Figura 18). A conservação dessa arquitetura de domínios associada as peculiaridades na região de dimerização, reforça a importância da proteína para vias do c-di-GMP.

Foram feitos alinhamentos de algumas estruturas resolvidas de EALs, através do programa Pymol, das sequências de EALs cataliticamente competentes provindos de TBD1265⁸², BlrP1¹¹⁵, STM3615¹⁶ e MorA (apo e ligada)¹¹⁵, da receptora de c-di-GMP LapD¹² e da proteína inativa *in vitro* YkuL.¹⁰⁶ Nesses alinhamentos foi constatado que STM3615 apresenta uma hélice $\alpha 5$ compatível com a forma reprimida da hélice $\alpha 5$ de MorA, que ocorre

na forma apo (na forma reprimida, a hélice $\alpha 5$ acaba englobando resíduos vizinhos do sítio de dimerização YS).

Figura 87 - Região do alinhamento envolvida na dimerização de alguns domínios EALs.

A numeração mostrada acima do alinhamento é a sequência da referência TDB1265. As regiões em vermelho destacam resíduos importantes para atividade ou para dimerização. As barras da primeira linha indicam a abrangência das hélices analisadas, tomada como referência a estrutura de TDB1265 (PDB:3n3t). As barras da segunda hélice indicam a abrangência dessas hélices encontradas em STM3615 (estrutura não publicada). A hélice $\alpha 4$ está logo após ao E619, que coordena o loop nas estruturas diméricas. As hélices $\alpha 5$ e $\alpha 6$ compõe a interface de dimerização, junto com o motivo DDFG(T/A)GYS.

Fonte: Elaborada pelo autor.

Em STM3615, a hélice $\alpha 5$ se apresenta deslocada em relação a posição observada em outras EALs usadas para comparação. Resíduos do loop 5 em STM3615 se apresentam em uma conformação particular, afastada do sítio ativo do EAL, que mais se aproxima do que ocorre com na forma inativa de MorA. Com o programa disponibilizado *online* PIC (dados não mostrados) foi possível detectar interações do loop 4 com o loop 5. Da estrutura de STM3615, o loop 4 se encontra mais estendido e próximo ao loop 5 (Figura 88). O deslocamento do loop 5 pode ajudar a explicar porque STM3615 possui uma conformação naturalmente favorável para formação de dímeros, independente do substrato c-di-GMP.

A não conservação da região de dimerização do EAL encontrada em STM3615 provavelmente pode ser um dos indicativos da presença de outros mecanismos para o controle da atividade PDE. Para proteína RbdA, por exemplo, a forma dimérica ocorre entre e a hélice conectora entre PAS e GGDEF e $\alpha 6'$ do domínio EAL (PDB: 5xge, Figura 26A). Essa dimerização pode explicar a não conservação do motivo regulatório, que tem como sequência DDFCAGMSS.¹¹⁹ Para comprovação das hipóteses apresentadas, dados experimentais devem ser inclusos nas análises.

Figura 88 - Estruturas das regiões de dimerização de alguns domínios EALs com estrutura resolvida.

Em destaque nas estruturas estão as regiões envolvidas com a dimerização de EALs. Dos dímeros mostrados, uma das unidades está em laranja e a outra unidade está em marrom. As estruturas foram alinhadas e editadas pelo programa Pymol.

Fonte: Elaborada pelo autor.

4 CONCLUSÕES E PERSPECTIVAS

Estudando domínios isolados de PA0847 foi possível observar que o domínio GGDEF promove o estado dimérico de um equilíbrio monômero-dímero da porção citoplasmática de PA0847. A interface não pode ser a mesma observada na forma inibida I, dentre outros, o GGDEF isolado seria observado com o ligante copurificado. A dimerização de um domínio GGDEF, sem c-di-GMP, é inusitado, dado que outros casos o domínio é observado em formas monoméricas. Portanto, isso compõe uma das novidades do estudo.

Concluiu-se que na fração citoplasmática de PA0847 ocorre uma segunda forma de dimerização pelo domínio HAMP, que permite a observação de uma parcela de tetrâmeros. Essa forma de dimerização foi indicada como a responsável por promover dímeros ativos de PA0847. Isso porque foi associada a mutação AXXD (com a construção HPG-1^{AXXD}) com o impedimento da dimerização pelo domínio GGDEF, que não eliminou a atividade.

Outras construções de PA0847 com fragmentos do domínio HAMP, como HPG-2, HPG-3 e PG, também devem formar um tetrâmero, mas de maneira transiente, por isso não observado em solução. Isso foi usado para explicar a menor atividade dessas construções é como as construções que forma dímeros dependentes do RXXD podem se tornar ativas. Além disso, a formação de tetrâmeros exige uma menor mudança conformacional das unidades para a obtenção de dímero ativo.

Caso for comprovada a relevância fisiológica da regulação um sistema de sinalização por tetrâmeros do domínio HAMP, os resultados com PA0847 seriam um caso inédito na literatura, dado que em outros exemplos descritos a forma funcional de HAMPs é a dimérica. Futuros experimentos podem avaliar se os enovelamentos conhecidos para o domínio HAMP, HAMP(A) e HAMP(B), seriam válidos para PA0847 e em caso positivo como esses enovelamentos estariam participando da regulação da proteína. Um possível experimento, por exemplo, é a substituição do domínio HAMP de PA0847 por um domínio HAMP de outra proteína, que apresenta a conformação HAMP(A) ou que apresente o enovelamento HAMP(B).

A adição do domínio PAS na construção GGDEF foi suficiente para proteína obter atividade. Foi possível supor que o domínio estaria participando de alguma forma na ligação ao c-di-GMP, direta ou indiretamente, dado que ocorre uma interface de interação entre PAS e GGDEF e ainda sim a adição gradual de elementos do domínio HAMP diminuiu gradativamente a tendência da construção copurificar com c-di-GMP. Possíveis próximos passos seriam investigar as hélices conectoras J α e Ncap e como elas poderiam atuar na

sinalização. Exemplos de experimentos com essas hélices é a simples análise de resíduos hidrofóbicos na busca de padrões de *coiled-coils*, ou a mutação ou deleção de resíduos dessas hélices conectoras para testar hipóteses.

Mais estudos seriam necessários para concluir o papel do domínio PAS na autoinibição de PA0847. Por exemplo, deve existir uma arginina como papel de sítio I(s) no domínio PAS para ligação ao c-di-GMP, como visto outros casos de heterodímeros intermediados por c-di-GMP (seção 1.8.2, Figura 14). Para buscar essa arginina, pode-se realizar triagem de mutações sítio-dirigidas em diferentes resíduos.

Resultados de cristalografia envolvendo domínios de PA0847 poderiam auxiliar nos detalhes moleculares dos possíveis estados ativos e inativos de PA0847, provavelmente guiados pelo domínio HAMP. Uma das questões que poderiam ser esclarecidas com a estrutura é sobre as formas mais alongadas e globulares das construções de PA0847, com e sem c-di-GMP (seção 3.1.4.1 e 3.1.4.2). Essas formas podem estar relacionadas inclusive com as diferentes respostas do domínio HAMP.

Nesse sentido, outras estratégias poderiam ser empregadas para obtenção de cristais, como por exemplo, explorar os equilíbrios oligoméricos, seja por mudanças no tampão, ou através de mutações favorecessem uma única forma oligomérica. Além disso, outras técnicas como EPR (ressonância paramagnética eletrônica), RMN (ressonância magnética nuclear) ou SAXS (espalhamento de raios X a baixo ângulo) poderiam ser exploradas para obter dados sobre a dinâmica dos domínios de PA0847.

Resultados de *pull-down*, duplo híbrido e termoforese confirmaram a interação entre PA0847 e HsbD e mostraram que ela é dependente da presença de todos os domínios de PA0847. Outros fatores ainda precisam ser comprovados, por exemplo, se a interação depende ou não da formação de tetrâmeros ou se existe uma interação direta entre HsbD-Nt e o domínio GGDEF. A conservação do padrão de HsbD-Nt em diferentes organismos e em proteínas com arquitetura de domínios semelhantes indica um sistema conservado de sinalização e ressalta a importância dessa região. Outros resultados experimentais poderiam auxiliar a explicar a função dessa interação, tais como experimentos em células de *P. aeruginosa*. Análises adicionais *in vitro* poderiam auxiliar a investigar como PA0847 poderia integrar a função originalmente descrita para HsbD, por exemplo, na localização de HsbD no polo celular. Outros experimentos poderiam investigar se, de fato, o domínio CHASE4 periplasmático de PA0847 está relacionada com a osmorregulação do sulfato.

Das análises de correlação, não foi possível encontrar uma comunidade que incluísse resíduos do GGDEF e EAL ativo. Essa comunidade poderia, por exemplo, indicar algum aspecto sobre casos de GGDEF e EAL bifuncionais ou sobre a transmissão de mudança conformacional entre esses domínios. Outra comunidade que se esperava encontrar é alguma que explicasse a dimerização constitutiva do EAL de STM3615. Entretanto, outros resultados de comunidades trouxeram potenciais aspectos ainda não explorados para proteínas com GGDEF e EAL em tandem. Dando exemplos, as comunidades 1 e comunidade 2 se destacaram pelo número de resíduos e por envolverem aqueles importantes para a atividade e para autorregulação (sítio I) do domínio GGDEF, respectivamente. Foi possível fazer essa relação devido à presença de resíduos conhecidos como importantes para essas funções. Dessas comunidades os resíduos Y439, da comunidade 1, e T363, da comunidade 2, se destacaram pelo número elevado de correlações e por não possuírem nenhuma descrição na literatura. Esses dois resíduos podem, portanto, revelar novos aspectos sobre a regulação da atividade DGC.

As comunidades 3, 4 e 5, por serem mais frequentes em sequências com o GGDEF e o EAL conservado, podem ter função regulatória para as enzimas bifuncionais do c-di-GMP. Uma vez que essas comunidades envolvem resíduos em regiões estratégicas no domínio, expostas ao solvente e envolvendo estruturas e resíduos importantes para atividade e para o reconhecimento do c-di-GMP.

Com os dados de correlação seria ideal novas análises para refinar os resultados. Uma dessas análises adicionais poderia ser a aplicação de outros métodos de correlação, como DCA (*Direct Coupling Analysis*).²⁴⁶ Novas análises poderiam ser feitas ainda classificando as sequências de acordo com a presença de regiões preditas transmembranares. Outra opção é realizar análises com número maior de sequências, especialmente daquelas que possuem apenas o GGDEF ativo ou os dois domínios degenerados.

Baseadas nos dados observados de proteínas resolvidas foi possível observar uma interface particular em STM3615, presente em proteínas semelhantes de outros organismos. Não foi possível correlacionar diretamente uma das comunidades das análises de correlação com a característica peculiar de STM3615, com EAL intrinsecamente dimérico.

Com as análises estruturais, foi sugerido que as alterações na interface de dimerização presente em STM3615 e proteínas correlatas, poderiam ser responsáveis por conferir dimerização constitutiva de STM3615. tal como o posicionamento do loop 5, por meio de interações com o loop 4 e a modificação das hélices $\alpha 5$ e $\alpha 6'$. A confecção de mutantes para

essas regiões poderia desestabilizar o dímero ou mesmo eliminar a atividade da STM3615, e isso pode auxiliar na comprovação da hipótese sobre a interface particular de STM3615.

REFERÊNCIAS

- 1 JENAL, U.; REINDERS, A.; LORI, C. Cyclic di-GMP: second messenger extraordinaire. **Nature Reviews Microbiology**, v. 15, n. 5, p. 271-284, May 2017.
- 2 MCDOUGALD, D. et al. Should we stay or should we go: mechanisms and ecological consequences for biofilm dispersal. **Nature Reviews Microbiology**, v. 10, n. 1, p. 39-50, Jan. 2012.
- 3 ROMLING, U.; LIANG, Z. X.; DOW, J. M. Progress in understanding the molecular basis underlying functional diversification of cyclic dinucleotide turnover proteins. **Journal of Bacteriology**, v. 199, n. 5, Mar. 2017.
- 4 MOUGEL, C.; ZHULIN, I. B. CHASE: an extracellular sensing domain common to transmembrane receptors from prokaryotes, lower eukaryotes and plants. **Trends in Biochemical Sciences**, v. 26, n. 10, p., 582, Oct. 2001.
- 5 ANANTHARAMAN, V.; ARAVIND, L. The CHASE domain: a predicted ligand-binding module in plant cytokinin receptors and other eukaryotic and bacterial receptors. **Trends in Biochemical Sciences**, v. 26, n. 10, p. 579, Oct. 2001.
- 6 ZHULIN, I. B.; NIKOLSKAYA, A. N.; GALPERIN, M. Y. Common extracellular sensory domains in transmembrane receptors for diverse signal transduction pathways in Bacteria and Archaea. **Journal of Bacteriology**, v. 185, n. 1, p. 285, Jan. 2003.
- 7 HENGGE, R. et al. Systematic nomenclature for GGDEF and EAL domain-containing cyclic Di-GMP turnover proteins of escherichia coli. **Journal of Bacteriology**, v. 198, n. 1, p. 7-11, Jan. 2016.
- 8 VALENTINI, M. et al. The diguanylate cyclase HsbD intersects with the HptB regulatory cascade to control pseudomonas aeruginosa biofilm and motility. **Plos Genetics**, v. 12, n. 10, Oct. 2016. doi.org/10.1371/journal.pgen.1006354.
- 9 DAHLSTROM, K. M. et al. A multimodal strategy used by a large c-di-GMP network. **Journal of Bacteriology**, v. 200, n. 8, Apr. 2018. doi: 10.1128/JB.00703-17.
- 10 HENGGE, R. Trigger phosphodiesterases as a novel class of c-di-GMP effector proteins. **Philosophical Transactions of the Royal Society B-biological sciences**, v. 371, n. 1707, Nov. 2016. doi: 10.1098/rstb.2015.0498.

11 MASSIE, J. P. et al. Quantification of high-specificity cyclic diguanylate signaling. **Proceedings of the National Academy of Sciences of the United States of America**, v. 109, n. 31, p. 12746-12751, July 2012.

12 DAHLSTROM, K. M. et al. Contribution of physical interactions to signaling specificity between a diguanylate cyclase and its effector. **Mbio**, v. 6, n. 6, Nov-Dec. 2015. doi: 10.1128/mBio.01978-15.

13 CARDOSO, A. R. **Mapeamento global de interações proteicas nas vias de sinalização mediadas por c-di-GMP de *Pseudomonas aeruginosa***. 2016. 107 p. Dissertação (Mestrado em Física Aplicada) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2016.

14 DEXTER, J. P.; GUNAWARDENA, J. Dimerization and bifunctionality confer robustness to the isocitrate dehydrogenase regulatory system in *Escherichia coli*. **Journal of Biological Chemistry**, v. 288, n. 8, p. 5770-5778, Feb. 2013.

15 DEXTER, J. P.; DASGUPTA, T.; GUNAWARDENA, J. Invariants reveal multiple forms of robustness in bifunctional enzyme systems. **Integrative Biology**, v. 7, n. 8, p. 883-894, 2015.

16 ROSSETO, F. R. **Estudos estruturais e funcionais de STM3615 de *Salmonella enterica*: uma proteína contendo ambos os domínios GGDEF-EAL envolvidos na biossíntese de c-di-GMP**. 2016. 95 p. Tese (Doutorado em Ciências) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2016.

17 OPPENHEIMER-SHAANAN, Y.; STEINBERG, N.; KOLODKIN-GAL, I. Small molecules are natural triggers for the disassembly of biofilms. **Trends in Microbiology**, v. 21, n. 11, p. 594-601, Nov. 2013.

18 SCHWEINITZER, T.; JOSEPHANS, C. Bacterial energy taxis: a global strategy? **Archives of Microbiology**, v. 192, n. 7, p. 507-520, Apr. 2010.

19 ROGGO, C.; VAN DER MEER, J. R. Miniaturized and integrated whole cell living bacterial sensors in field applicable autonomous devices. **Current Opinion in Biotechnology**, v. 45, p. 24-33, June 2017. doi.org/10.1016/j.copbio.2016.11.023.

20 RAVIKUMAR, S. et al. Engineered microbial biosensors based on bacterial two-component systems as synthetic biotechnology platforms in bioremediation and biorefinery. **Microbial Cell Factories**, v. 16, p. 62, Apr. 2017. doi.org/10.1186/s12934-017-0675-z.

21 ZHANG, F. Z.; KEASLING, J. Biosensors and their applications in microbial metabolic engineering. **Trends in Microbiology**, v. 19, n. 7, p. 323-329, July 2011.

22 KOH, C. L. et al. Plant-derived natural products as sources of anti-quorum sensing compounds. **Sensors**, v. 13, n. 5, p. 6217-6228, maio 2013.

23 SCHACHTER, B. Slimy business-the biotechnology of biofilms. **Nature Biotechnology**, v. 21, n. 4, p. 361-365, Apr. 2003.

24 HA, D. G.; O'TOOLE, G. A. c-di-GMP and its effects on biofilm formation and dispersion: a pseudomonas aeruginosa review. **Microbiology Spectrum**, v. 3, n. 2, Apr. 2015. doi: 10.1128/microbiolspec.MB-0003-2014.

25 GALLOWAY, W. et al. Applications of small molecule activators and inhibitors of quorum sensing in Gram-negative bacteria. **Trends in Microbiology**, v. 20, n. 9, p. 449-458, Sept. 2012.

26 RICKARD, A. H. et al. Bacterial coaggregation: an integral process in the development of multi-species biofilms. **Trends in Microbiology**, v. 11, n. 2, p. 94-100, Feb. 2003.

27 COSTERTON, J. W.; STEWART, P. S. Battling biofilms - the war is against bacterial colonies that cause some of the most tenacious infections known. the weapon is knowledge of the enemy's communication system. **Scientific American**, v. 285, n. 1, p. 75-81, July 2001.

28 MONROE, D. Looking for chinks in the armor of bacterial biofilms. **Plos Biology**, v. 5, n. 11, p. 2458-2461, Nov. 2007.

29 MARTIN-RODRIGUEZ, A. J.; ROMLING, U. Nucleotide second messenger signaling as a target for the control of bacterial biofilm formation. **Current Topics in Medicinal Chemistry**, v. 17, n. 17, p. 1928-1944, 2017.

30 WIGGERS, H. J. et al. Identification of anti-inflammatory and anti- hypertensive drugs as inhibitors of bacterial diguanylate cyclases. **Journal of the Brazilian Chemical Society**, v. 29, n. 2, p. 297-309, Feb. 2018.

31 RYAN, R. P. Cyclic di-GMP signalling and the regulation of bacterial virulence. **Microbiology-SGM**, v. 159, p. 1286-1297, July 2013.

32 FLEMMING, H. C.; WINGENDER, J. The biofilm matrix. **Nature Reviews Microbiology**, v. 8, n. 9, p. 623-633, Sept. 2010

33 FLEMMING, H.C.; NEU, T. R.; WOZNIAK, D. J. The EPS matrix: The "House of Biofilm cells". **Journal of Bacteriology**, v. 189, n. 22, p. 7945-7947, Nov. 2007.

34 KARATAN, E.; WATNICK, P. Signals, regulatory networks, and materials that build and break bacterial biofilms. **Microbiology and Molecular Biology Reviews**, v. 73, n. 2, p. 310, June 2009.

35 HALL-STOODLEY, L.; STOODLEY, P. Evolving concepts in biofilm infections. **Cellular Microbiology**, v. 11, n. 7, p. 1034-1043, July 2009.

36 UPPULURI, P.; LOPEZ-RIBOT, J. L. Go forth and colonize: dispersal from clinically important microbial biofilms. **Plos Pathogens**, v. 12, n. 2, Feb. 2016.
doi.org/10.1371/journal.ppat.1005397.

37 DAVIES, D. Understanding biofilm resistance to antibacterial agents. **Nature Reviews Drug Discovery**, v. 2, n. 2, p. 114-122, Feb. 2003.

38 COSTERTON, J. W.; STEWART, P. S.; GREENBERG, E. P. Bacterial biofilms: a common cause of persistent infections. **Science**, v. 284, n. 5418, p. 1318-1322, May 1999.

39 COSTERTON, J. W.; STEWART, P. S. Battling biofilms. **Scientific American**, v. 285, n. 1, p. 75-81, July 2001.

40 SREY, S.; JAHID, I. K.; HA, S.-D. Biofilm formation in food industries: a food safety concern. **Food Control**, v. 31, n. 2, p. 572-585, June 2013.

41 PREVOTEAU, A.; RABAEY, K. Electroactive biofilms for sensing: reflections and perspectives. **ACS Sensors**, v. 2, n. 8, p. 1072-1085, Aug. 2017.

42 AGOSTINO, V. et al. Electrochemical analysis of microbial fuel cells based on enriched biofilm communities from freshwater sediment. **Electrochimica Acta**, v. 237, p. 133-143, May 2017. doi.org/10.1016/j.electacta.2017.03.186.

43 SINGH, R.; PAUL, D.; JAIN, R. K. Biofilms: implications in bioremediation. **Trends in Microbiology**, v. 14, n. 9, p. 389-397, Sept. 2006.

44 LEBEAUX, D. et al. From in vitro to in vivo Models of Bacterial Biofilm-Related Infections. **Pathogens**, v. 2, n. 2, p. 68, 2013.

45 BRYERS, J. D. Medical biofilms. **Biotechnology and Bioengineering**, v. 100, n. 1, p. 1-18, May 2008.

- 46 DAROUICHE, R. O. Current concepts - treatment of infections associated with surgical implants. **New England Journal of Medicine**, v. 350, n. 14, p. 1422-1429, Apr. 2004.
- 47 JAMAL, M. et al. Bacterial biofilm and associated infections. **Journal of the Chinese Medical Association**, v. 81, n. 1, p. 7-11, Jan. 2018.
- 48 JOO, H. S.; OTTO, M. Molecular basis of in vivo biofilm formation by bacterial pathogens. **Chemistry & Biology**, v. 19, n. 12, p. 1503-1513, Dec. 2012.
- 49 LYCZAK, J. B.; CANNON, C. L.; PIER, G. B. Lung infections associated with cystic fibrosis. **Clinical Microbiology Reviews**, v. 15, n. 2, p. 194, Apr. 2002.
- 50 ORGANS FOR LIFE. **Cystic Fibrosis**. 2014. Disponível em: <<http://organsforlife.org.uk/cystic-fibrosis/>>. Acesso em: 13 jul. 2018.
- 51 SCIENCE PHOTO LIBRARY. **Researchers develop first human model of P. aeruginosa infection in cystic fibrosis lungs**. health-innovations.org, 2015. Disponível em: <<https://health-innovations.org/2015/03/18/researchers-develop-first-human-model-of-p-aeruginosa-infection-in-cystic-fibrosis-lungs/>>. Acesso em: 13 jul. 2018.
- 52 ATKINS, L. **Salmonella outbreak could spread to hundreds more**. 2010. Disponível em: <<https://shop2insure.wordpress.com/2010/08/18/331/>>. Acesso em: 13 jul. 2018.
- 53 BERGAMIN, F. **Teamwork against mutant free riders**. 2013. Disponível em: <<https://phys.org/news/2013-02-teamwork-mutant-free-riders.html>>. Acesso em: 13 jul. 2018.
- 54 BENNETT, W. M. O. L. D. A. S. **WHO publishes list of bacteria for which new antibiotics are urgently needed**. Geneva, 2017. Disponível em: <<http://www.who.int/en/news-room/detail/27-02-2017-who-publishes-list-of-bacteria-for-which-new-antibiotics-are-urgently-needed>>. Acesso em: 13 jul. 2018.
- 55 RAMSEY, M. M.; WHITELEY, M. Pseudomonas aeruginosa attachment and biofilm development in dynamic environments. **Molecular Microbiology**, v. 53, n. 4, p. 1075-1087, Aug. 2004.
- 56 WALKER, T. S. et al. Pseudomonas aeruginosa-plant root interactions. Pathogenicity, biofilm formation, and root exudation. **Plant Physiology**, v. 134, n. 1, p. 320-331, Jan. 2004.
- 57 FABREGA, A.; VILA, J. Salmonella enterica serovar typhimurium skills to succeed in the host: virulence and regulation. **Clinical Microbiology Reviews**, v. 26, n. 2, p. 308-341, Apr. 2013.

58 MAURICE, D. H. et al. Advances in targeting cyclic nucleotide phosphodiesterases. **Nature Reviews Drug Discovery**, v. 13, n. 4, p. 290-314, Apr. 2014.

59 HENGGE, R. et al. Bacterial signal transduction by cyclic di-gmp and other nucleotide second messengers. **Journal of Bacteriology**, v. 198, n. 1, p. 15-26, Jan. 2016.

60 RAO, F. et al. The functional role of a conserved loop in EAL domain-based cyclic di-GMP-specific phosphodiesterase. **Journal of Bacteriology**, v. 191, n. 15, p. 4722-4731, Aug. 2009.

61 ROSS, P. et al. Regulation of cellulose synthesis in acetobacter-xylinum by cyclic diguanylic acid. **Nature**, v. 325, n. 6101, p. 279-281, Jan. 1987.

62 RALL, T. W.; SUTHERLAND, E. W.; BERTHET, J. THE relationship of epinephrine and glucagon to liver phosphorylase .4. effect of epinephrine and glucagon on the reactivation of phosphorylase in liver homogenates. **Journal of Biological Chemistry**, v. 224, n. 1, p. 463-475, 1957.

63 SUTHERLAND, E. W.; RALL, T. W. Fractionation and characterization of a cyclic adenine ribonucleotide formed by tissue particles. **Journal of Biological Chemistry**, v. 232, n. 2, p. 1077-1091, 1958.

64 MAKMAN, R. S.; SUTHERLAND, E. W. Presence of cyclic 3',5'-Adenosine phosphate (CA) IN E COLI. **Federation Proceedings**, v. 22, n. 2, p. 470, 1963.

65 OKABAYASHI, T.; YOSHIMOTO, A.; IDE, M. Excretion of adenosine-3' ,5'-phosphate in culture broth of brevibacterium liquefaciens. **Archives of Biochemistry and Biophysics**, v. 100, n. 1, p. 158-&, 1963.

66 ASHMAN, D. F. et al. Isolation of adenosine 3,5-monophosphate and guanosine 3,5-monophosphate from rat urine. **Biochemical and Biophysical Research Communications**, v. 11, n. 4, p. 330-&, 1963.

67 BERNLOHR, R. W.; HADDOX, M. K.; GOLDBERG, N. D. Cyclic guanosine 3'-5'-monophosphate in escherichia-coli and bacillus-lichenformis. **Journal of Biological Chemistry**, v. 249, n. 13, p. 4329-4331, 1974.

68 LINDER, J. U. cGMP production in bacteria. **Molecular and Cellular Biochemistry**, v. 334, n. 1-2, p. 215-219, Jan. 2010.

69 PIERCE, K. L.; PREMONT, R. T.; LEFKOWITZ, R. J. Seven-transmembrane receptors. **Nature Reviews Molecular Cell Biology**, v. 3, n. 9, p. 639-650, Sept. 2002.

70 BERRIDGE, M. J.; LIPP, P.; BOOTMAN, M. D. The versatility and universality of calcium signalling. **Nature Reviews Molecular Cell Biology**, v. 1, n. 1, p. 11-21, Oct. 2000.

71 CASHEL, M. The control of ribonucleic acid synthesis in Escherichia coli IV. Relevance of unusual phosphorylated compounds from amino acid-starved stringent strains. **Journal of Biological Chemistry**, v. 244, n. 12, p. 8, 1969.

72 DALEBROUX, Z. D.; SWANSON, M. S. ppGpp: magic beyond RNA polymerase. **Nature Reviews Microbiology**, v. 10, n. 3, p. 203-212, Mar. 2012.

73 CORRIGAN, R. M.; GRUENDLING, A. Cyclic di-AMP: another second messenger enters the fray. **Nature Reviews Microbiology**, v. 11, n. 8, p. 513-524, Aug. 2013.

74 DAVIES, B. W. et al. Coordinated regulation of accessory genetic elements produces cyclic di-nucleotides for V. cholerae virulence. **Cell**, v. 149, n. 2, p. 358-370, Apr. 2012.

75 WU, J. X. et al. Cyclic GMP-AMP Is an endogenous second messenger in innate immune signaling by cytosolic DNA. **Science**, v. 339, n. 6121, p. 826-830, Feb. 2013.

76 SUN, L. J. et al. Cyclic GMP-AMP Synthase is a Cytosolic DNA sensor that activates the type I interferon pathway. **Science**, v. 339, n. 6121, p. 786-791, Feb. 2013.

77 ABLASSER, A. et al. cGAS produces a 2'-5'-linked cyclic dinucleotide second messenger that activates STING. **Nature**, v. 498, n. 7454, p. 380-+, June 2013.

78 TAO, J. L.; ZHOU, X.; JIANG, Z. F. cGAS-cGAMP-STING: The three musketeers of Cytosolic DNA sensing and signaling. **Imbmb Life**, v. 68, n. 11, p. 858-870, Nov. 2016.

79 WOODWARD, J. J.; IAVARONE, A. T.; PORTNOY, D. A. c-di-AMP secreted by intracellular listeria monocytogenes activates a host type I interferon response. **Science**, v. 328, n. 5986, p. 1703-1705, June 2010.

80 BURDETTE, D. L. et al. STING is a direct innate immune sensor of cyclic di-GMP. **Nature**, v. 478, n. 7370, p. 515-U111, Oct. 27 2011.

81 SIMM, R. et al. Quantitative determination of cyclic diguanosine monophosphate concentrations in nucleotide extracts of bacteria by matrix-assisted laser

desorption/ionization-time-of-flight mass spectrometry. **Analytical Biochemistry**, v. 386, n. 1, p. 53-58, Mar. 2009.

82 TCHIGVINTSEV, A. et al. Structural insight into the mechanism of c-di-GMP hydrolysis by EAL domain phosphodiesterases. **Journal of Molecular Biology**, v. 402, n. 3, p. 524-538, Sept. 2010.

83 RAO, F. et al. Catalytic mechanism of cyclic Di-GMP-specific phosphodiesterase: a study of the EAL domain-containing RocR from *Pseudomonas aeruginosa*. **Journal of Bacteriology**, v. 190, n. 10, p. 3622-3631, May 2008.

84 HYODO, M.; SATO, Y.; HAYAKAWA, Y. Synthesis of cyclic bis(3'-5')diguanlylic acid (c-di-GMP) analogs. **Tetrahedron**, v. 62, n. 13, p. 3089-3094, Mar. 2006.

85 ZHRINGER, F.; MASSA, C.; SCHIRMER, T. Efficient enzymatic production of the bacterial second messenger c-di-GMP by the diguanylate Cyclase YdeH from *E. coli*. **Applied Biochemistry and Biotechnology**, v. 163, n. 1, p. 71-79, Jan. 2011.

86 WANG, J. X. et al. Conservative change to the phosphate moiety of cyclic diguanlylic monophosphate remarkably affects its polymorphism and ability to bind DGC, PDE, and PilZ proteins. **Journal of the American Chemical Society**, v. 133, n. 24, p. 9320-9330, June 2011.

87 ZHANG, Z. Y. et al. Polymorphism of the signaling molecule c-di-GMP. **Journal of the American Chemical Society**, v. 128, n. 21, p. 7015-7024, May 2006.

88 CHOU, S. H.; GALPERIN, M. Y. Diversity of Cyclic Di-GMP-binding proteins and mechanisms. **Journal of Bacteriology**, v. 198, n. 1, p. 32-46, Jan. 2016.

89 KRASTEVA, P. V.; SONDERMANN, H. Versatile modes of cellular regulation via cyclic dinucleotides. **Nature Chemical Biology**, v. 13, n. 4, p. 350-359, Apr. 2017.

90 SUNDRIYAL, A. et al. Inherent regulation of EAL domain-catalyzed hydrolysis of second messenger Cyclic di-GMP. **Journal of Biological Chemistry**, v. 289, n. 10, p. 6978-6990, Mar. 2014.

91 AMIKAM, D.; GALPERIN, M. Y. PilZ domain is part of the bacterial c-di-GMP binding protein. **Bioinformatics**, v. 22, n. 1, p. 3-6, Jan. 2006.

92 FANG, X. et al. GIL, a new c-di-GMP-binding protein domain involved in regulation of cellulose synthesis in enterobacteria. **Molecular Microbiology**, v. 93, n. 3, p. 439-452, Aug. 2014.

- 93 MERRITT, J. H. et al. Specific control of pseudomonas aeruginosa surface-associated behaviors by two c-di-GMP diguanylate Cyclases. **Mbio**, v. 1, n. 4, Sep-Oct. 2010.
- 94 CHAN, C. et al. Structural basis of activity and allosteric control of diguanylate cyclase. **Proceedings of the National Academy of Sciences of the United States of America**, v. 101, n. 49, p. 17084-17089, Dec. 2004.
- 95 WASSMANN, P. et al. Structure of BeF₃--modified response regulator PleD: implications for diguanylate cyclase activation, catalysis, and feedback inhibition. **Structure**, v. 15, n. 8, p. 915-927, Aug. 2007.
- 96 ZAHRINGER, F. et al. Structure and signaling mechanism of a zinc-sensory diguanylate cyclase. **Structure**, v. 21, n. 7, p. 1149-1157, July 2013.
- 97 SCHIRMER, T. C-di-GMP synthesis: structural aspects of evolution, catalysis and regulation. **Journal of Molecular Biology**, v. 428, n. 19, p. 3683-3701, Sept. 2016.
- 98 HUNTER, J. L. et al. The Vibrio cholerae diguanylate cyclase VCA0965 has an AGDEF active site and synthesizes cyclic di-GMP. **BMC Microbiology**, v. 14, Feb. 2014. doi: 10.1186/1471-2180-14-22.
- 99 PEREZ-MENDOZA, D. et al. A multi-repeat adhesin of the phytopathogen, Pectobacterium atrosepticum, is secreted by a Type I pathway and is subject to complex regulation involving a non-canonical diguanylate cyclase. **Molecular Microbiology**, v. 82, n. 3, p. 719-733, Nov. 2011.
- 100 MALONE, J. G. et al. The structure-function relationship of WspR, a Pseudomonas fluorescens response regulator with a GGDEF output domain. **Microbiology-SGM**, v. 153, Pt. 4, p. 980-994, Apr. 2007.
- 101 KIRILLINA, O. et al. HmsP, a putative phosphodiesterase, and HmsT, a putative diguanylate cyclase, control Hms-dependent biofilm formation in Yersinia pestis. **Molecular Microbiology**, v. 54, n. 1, p. 75-88, Oct. 2004.
- 102 GALPERIN, M. Y.; NIKOLSKAYA, A. N.; KOONIN, E. V. Novel domains of the prokaryotic two-component signal transduction systems. **FEMS Microbiology Letters**, v. 203, n. 1, p. 11-21, Sept. 2001.
- 103 DE, N. et al. Phosphorylation-independent regulation of the diguanylate cyclase WspR. **Plos Biology**, v. 6, n. 3, p. 601-617, Mar. 2008.

- 104 MANIKANDAN, K. et al. Two-step synthesis and hydrolysis of cyclic di-AMP in mycobacterium tuberculosis. **Plos One**, v. 9, n. 1, p. e86096, Jan. 2014.
- 105 PEI, J. M.; GRISHIN, N. V. GGDEF domain is homologous to adenylyl cyclase. **Proteins-structure function and genetics**, v. 42, n. 2, p. 210-216, Feb. 2001.
- 106 MINASOV, G. et al. Crystal structures of YkuI and Its complex with second messenger cyclic Di-GMP suggest catalytic mechanism of phosphodiester bond cleavage by EAL domains. **Journal of Biological Chemistry**, v. 284, n. 19, p. 13174-13184, May 2009.
- 107 CHRISTEN, M. et al. Identification and characterization of a cyclic di-GMP-specific phosphodiesterase and its allosteric control by GTP. **Journal of Biological Chemistry**, v. 280, n. 35, p. 30829-30837, Sept. 2005.
- 108 LI, Z. et al. Structures of the PelD cyclic diguanylate effector involved in pellicle formation in pseudomonas aeruginosa PAO1. **Journal of Biological Chemistry**, v. 287, n. 36, p. 30191-30204, Aug. 2012.
- 109 NAVARRO, M. et al. Structural analysis of the GGDEF-EAL domain-containing c-di-GMP receptor FimX. **Structure**, v. 17, n. 8, p. 1104-1116, Aug. 2009.
- 110 SCHIRMER, T.; JENAL, U. Structural and mechanistic determinants of c-di-GMP signalling. **Nature Reviews Microbiology**, v. 7, n. 10, p. 724-735, Oct. 2009.
- 111 GIARDINA, G. et al. Investigating the allosteric regulation of YfiN from pseudomonas aeruginosa: clues from the structure of the catalytic domain. **PLoS ONE**, v. 8, n. 11, p. e81324, Nov. 2013.
- 112 DEEPTHI, A. et al. Structure of a diguanylate cyclase from thermotoga maritima: insights into activation, feedback inhibition and thermostability. **PLoS ONE**, v. 9, n. 10, Oct. 2014.
- 113 DE, N. et al. Determinants for the activation and autoinhibition of the diguanylate cyclase response regulator WspR. **Journal of Molecular Biology**, v. 393, n. 3, Oct. 2009.
- 114 COHEN, D. et al. Oligoribonuclease is a central feature of cyclic diguanylate signaling in Pseudomonas aeruginosa. **Proceedings of the National Academy of Sciences of the United States of America**, v. 112, n. 36, p. 11359-11364, Sept. 2015.
- 115 BARENDT, T. R. M. et al. Structure and mechanism of a bacterial light-regulated cyclic nucleotide phosphodiesterase. **Nature**, v. 459, n. 7249, p. 1015-U150, June 2009.

- 116 PHIPPEN, C. W. et al. Formation and dimerization of the phosphodiesterase active site of the *Pseudomonas aeruginosa* MorA, a bi-functional c-di-GMP regulator. **FEBS Letters**, v. 588, n. 24, p. 4631-4636, Dec. 2014.
- 117 ROMLING, U.; GALPERIN, M. Y.; GOMELSKY, M. Cyclic di-GMP: the first 25 Years of a universal bacterial second messenger. **Microbiology and Molecular Biology Reviews**, v. 77, n. 1, p. 1-52, Mar. 2013.
- 118 CHEN, M. W. et al. Structural insights into the regulatory mechanism of the response regulator rocR from *pseudomonas aeruginosa* in Cyclic Di-GMP signaling. **Journal of Bacteriology**, v. 194, n. 18, p. 4837-4846, Sept. 2012.
- 119 LIU, C. et al. Insights into biofilm dispersal regulation from the crystal structure of the PAS-GGDEF-EAL region of RbdA from *Pseudomonas aeruginosa*. **Journal of Bacteriology**, v. 200, n. 3, Feb. 2018. doi: 10.1128/JB.00515-17.
- 120 NAVARRO, M. V. A. S. et al. Structural basis for c-di-GMP-mediated inside-out signaling controlling periplasmic proteolysis. **PLoS Biology**, v. 9, n. 2, p. e1000588, Feb. 2011.
- 121 SESHASAYEE, A. S. N.; FRASER, G. M.; LUSCOMBE, N. M. Comparative genomics of cyclic-di-GMP signalling in bacteria: post-translational regulation and catalytic activity. **Nucleic Acids Research**, v. 38, n. 18, p. 5970-5981, Oct. 2010.
- 122 SARENKO, O. et al. More than enzymes that make or break cyclic Di-GMP-local signaling in the interactome of GGDEF/EAL domain proteins of *Escherichia coli*. **Mbio**, v. 8, n. 5, Sep-Oct. 2017. doi: 10.1128/mBio.01639-17.
- 123 GALPERIN, M. Y. A census of membrane-bound and intracellular signal transduction proteins in bacteria: bacterial IQ, extroverts and introverts. **BMC Microbiology**, v. 5, June 2005. doi.org/10.1186/1471-2180-5-35.
- 124 GALPERIN, M. Y.; GOMELSKY, M. Bacterial signal transduction modules: From genomics to biology. **ASM News**, v. 71, n. 7, p. 326-333, July 2005.
- 125 GALPERIN, M. Y. Structural classification of bacterial response regulators: diversity of output domains and domain combinations. **Journal of Bacteriology**, v. 188, n. 12, p. 4169-4182, June 2006.
- 126 PAO, G. M.; SAIER, M. H. Response regulators of bacterial signal-transduction systems - selective domain shuffling during evolution. **Journal of Molecular Evolution**, v. 40, n. 2, p. 136-154, Feb. 1995.

- 127 CHRISTEN, M. et al. Asymmetrical distribution of the second messenger c-di-GMP upon bacterial cell division. **Science**, v. 328, n. 5983, p. 1295-1297, June 2010.
- 128 AIROLA, M. V. et al. Architecture of the soluble receptor Aer2 indicates an In-Line mechanism for PAS and HAMP domain signaling. **Journal of Molecular Biology**, v. 425, n. 5, p. 886-901, Mar. 2013.
- 129 GAO, R.; STOCK, A. M. Biological insights from structures of two-component proteins. **Annual Review of Microbiology**, v. 63, p. 133-154, 2009. doi: 10.1146/annurev.micro.091208.073214.
- 130 PARKINSON, J. S. Signaling mechanisms of HAMP domains in chemoreceptors and sensor kinases. **Annual Review of Microbiology**, v. 64, p. 101-122, 2010. doi: 10.1146/annurev.micro.112408.134215.
- 131 DUNIN-HORKAWICZ, S.; LUPAS, A. N. Comprehensive analysis of HAMP domains: implications for transmembrane signal transduction. **Journal of Molecular Biology**, v. 397, n. 5, p. 1156-1174, Apr. 2010.
- 132 HE, K.; BAUER, C. E. Chemosensory signaling systems that control bacterial survival. **Trends in Microbiology**, v. 22, n. 7, p. 389-398, July 2014.
- 133 WADHAMS, G. H.; ARMITAGE, J. P. Making sense of it all: Bacterial chemotaxis. **Nature Reviews Molecular Cell Biology**, v. 5, n. 12, p. 1024-1037, Dec. 2004.
- 134 BRIEGEL, A. et al. Universal architecture of bacterial chemoreceptor arrays. **Proceedings of the National Academy of Sciences of the United States of America**, v. 106, n. 40, p. 17181-17186, Oct. 2009.
- 135 SOURJIK, V.; WINGREEN, N. S. Responding to chemical gradients: bacterial chemotaxis. **Current Opinion in Cell Biology**, v. 24, n. 2, p. 262-268, Apr. 2012.
- 136 UNIVESITY OF UTAH. Departament of Biology. Parkinson Lab. **An overview of E. coli chemotaxis**. Disponível em: http://chemotaxis.biology.utah.edu/Parkinson_Lab/projects/ecolichemotaxis/ecolichemotaxis.html >. Acesso em: 08 nov. 2015.
- 137 NEIDITCH, M. B. et al. Ligand-induced asymmetry in histidine sensor kinase complex regulates quorum sensing. **Cell**, v. 126, n. 6, p. 1095-1108, Sept. 2006.

138 PORTER, S. L.; WADHAMS, G. H.; ARMITAGE, J. P. Signal processing in complex chemotaxis pathways. **Nature Reviews Microbiology**, v. 9, n. 3, p. 153-165, Mar. 2011.

139 RUSSELL, M. H. et al. Integration of the second messenger c-di-GMP into the chemotactic signaling pathway. **Mbio**, v. 4, n. 2, Mar-Apr. 2013. doi: 10.1128/mBio.00001-13.

140 MOSCOSO, J. A. et al. The *Pseudomonas aeruginosa* sensor RetS switches type III and type VI secretion via c-di-GMP signalling. **Environmental Microbiology**, v. 13, n. 12, p. 3128-3138, Dec. 2011.

141 CLARKE, M. B. et al. The QseC sensor kinase: a bacterial adrenergic receptor. **Proceedings of the National Academy of Sciences of the United States of America**, v. 103, n. 27, p. 10420-10425, July 2006.

142 MASCHER, T.; HELMANN, J. D.; UNDEN, G. Stimulus perception in bacterial signal-transducing histidine kinases. **Microbiology and Molecular Biology Reviews**, v. 70, n. 4, p. 910-938, Dec. 2006.

143 CHEUNG, J.; HENDRICKSON, W. A. Sensor domains of two-component regulatory systems. **Current Opinion in Microbiology**, v. 13, n. 2, p. 116-123, Apr. 2010.

144 LACAL, J. et al. Sensing of environmental signals: classification of chemoreceptors according to the size of their ligand binding regions. **Environmental Microbiology**, v. 12, n. 11, p. 2873-2884, Nov. 2010.

145 UPADHYAY, A. A. et al. Cache domains that are homologous to, but different from PAS domains comprise the largest superfamily of extracellular sensors in prokaryotes. **PLoS Computational Biology**, v. 12, n. 4, Apr. 2016. doi: 10.1371/journal.pcbi.1004862.

146 ZHANG, Z.; HENDRICKSON, W. A. Structural characterization of the predominant family of histidine kinase sensor domains. **Journal of Molecular Biology**, v. 400, n. 3, p. 335-353, July 2010.

147 POKKULURI, P. R. et al. Structures and solution properties of two novel periplasmic sensor domains with c-type heme from chemotaxis proteins of *Geobacter sulfurreducens*: implications for signal transduction. **Journal of Molecular Biology**, v. 377, n. 5, p. 1498-1517, Apr. 2008.

148 BERNTSSON, R. P. A. et al. A structural classification of substrate-binding proteins. **Febs Letters**, v. 584, n. 12, p. 2606-2617, June 2010.

- 149 MOORE, J. O.; HENDRICKSON, W. A. An Asymmetry-to-symmetry switch in signal transmission by the histidine kinase receptor for TMAO. **Structure**, v. 20, n. 4, p. 729-741, Apr. 2012.
- 150 KRELL, T. et al. Bacterial sensor kinases: diversity in the recognition of environmental signals. **Annual Review of Microbiology**, v. 64, p. 539-559, 2010.
- 151 XU, M. et al. Structural studies of the periplasmic portion of the diguanylate cyclase CdgH from *Vibrio cholerae*. **Scientific Reports**, v. 7, May 2017. doi: 10.1038/s41598-017-01989-6
- 152 BADER, M. W. et al. Recognition of antimicrobial peptides by a bacterial sensor kinase. **Cell**, v. 122, n. 3, p. 461-472, Aug. 2005.
- 153 MOLNAR, K. S. et al. Cys-scanning disulfide cross linking and Bayesian modeling probe the transmembrane signaling mechanism of the histidine kinase, PhoQ. **Structure**, v. 22, n. 9, p. 1239-1251, Sept. 2014.
- 154 WOSTEN, M. et al. Function and regulation of the C4-dicarboxylate transporters in *Campylobacter jejuni*. **Frontiers in Microbiology**, v. 8, Feb. 2017. doi: 10.3389/fmicb.2017.00174.
- 155 HENRY, J. T.; CROSSON, S. Ligand-binding PAS domains in a genomic, cellular, and structural context. **Annual Review of Microbiology**, v. 65, p. 261-286, 2011.
- 156 NEIDITCH, M. B. et al. Regulation of LuxPQ receptor activity by the quorum-sensing signal autoinducer-2. **Molecular Cell**, v. 18, n. 5, p. 507-518, May 2005.
- 157 WARD, S. M.; BORMANS, A. F.; MANSON, M. D. Mutationally altered signal output in the narX (NarX-Tar) hybrid chemoreceptor. **Journal of Bacteriology**, v. 188, n. 11, p. 3944-3951, June 2006
- 158 CHEUNG, J.; LE-KHAC, M.; HENDRICKSON, W. A. Crystal structure of a histidine kinase sensor domain with similarity to periplasmic binding proteins. **Proteins-Structure Function and Bioinformatics**, v. 77, n. 1, p. 235-241, Oct. 2009.
- 159 MOEGLICH, A.; AYERS, R. A.; MOFFAT, K. Structure and signaling mechanism of Per-ARNT-Sim domains. **Structure**, v. 17, n. 10, p. 1282-1294, Oct. 2009.
- 160 GONG, W. M. et al. Structure of a biological oxygen sensor: a new mechanism for heme-driven signal transduction. **Proceedings of the National Academy of Sciences of the United States of America**, v. 95, n. 26, p. 15177-15182, Dec. 1998.

161 TAYLOR, B. L.; ZHULIN, I. B. PAS domains: internal sensors of oxygen, redox potential, and light. **Microbiology and Molecular Biology Reviews**, v. 63, n. 2, p. 479-+, June 1999.

162 BHATE, M. P. et al. Signal transduction in histidine kinases: insights from new structures. **Structure**, v. 23, n. 6, p. 981-994, June 2015.

163 HERROU, J.; CROSSON, S. Function, structure and mechanism of bacterial photosensory LOV proteins. **Nature Reviews Microbiology**, v. 9, n. 10, p. 713-723, Oct. 2011.

164 HEINTZ, U.; MEINHART, A.; WINKLER, A. Multi-PAS domain-mediated protein oligomerization of PpsR from *Rhodobacter sphaeroides*. **Acta Crystallographica D**, v. 70, p. 863-876, Mar. 2014.

165 WINNEN, B. et al. Role of the PAS sensor domains in the *Bacillus subtilis* sporulation kinase KinA. **Journal of Bacteriology**, v. 195, n. 10, p. 2349-2358, May 2013.

166 WANG, C. et al. Mechanistic insights revealed by the crystal structure of a histidine kinase with signal transducer and sensor domains. **Plos Biology**, v. 11, n. 2, Feb. 2013.

167 QI, Y. et al. A Flavin Cofactor-Binding PAS domain regulates c-di-GMP synthesis in AxDGC2 from *Acetobacter xylinum*. **Biochemistry**, v. 48, n. 43, p. 10275-10285, Nov. 2009.

168 KYNDT, J. A. et al. The photoactivated PYP domain of *Rhodospirillum rubrum* Ppr accelerates the recovery of the bacteriophytochrome domain after white light illumination. **Biochemistry**, v. 46, n. 28, p. 8256-8262, July 2007.

169 IMAMOTO, Y.; KATAOKA, M. Structure and photoreaction of photoactive yellow protein, a structural prototype of the PAS domain superfamily. **Photochemistry and Photobiology**, v. 83, n. 1, p. 40-49, Jan-Feb. 2007.

170 WATTS, K. J.; TAYLOR, B. L.; JOHNSON, M. S. PAS/poly-HAMP signalling in Aer-2, a soluble haem-based sensor. **Molecular Microbiology**, v. 79, n. 3, p. 686-699, Feb. 2011.

171 HO, Y. S. J.; BURDEN, L. M.; HURLEY, J. H. Structure of the GAF domain, a ubiquitous signaling motif and a new class of cyclic GMP receptor. **EMBO Journal**, v. 19, n. 20, p. 5288-5299, Oct. 2000.

172 KORYCINSKI, M. et al. STAC-a new domain associated with transmembrane solute transport and two-component signal transduction systems. **Journal of Molecular Biology**, v. 427, n. 20, p. 3327-3339, Oct. 2015.

173 ZORAGHI, R.; CORBIN, J. D.; FRANCIS, S. H. Properties and functions of GAF domains in cyclic nucleotide phosphodiesterases and other proteins. **Molecular Pharmacology**, v. 65, n. 2, p. 267-278, Feb. 2004.

174 BATEMAN, A. et al. UniProt: the universal protein knowledgebase. **Nucleic Acids Research**, v. 45, n. D1, p. D158-D169, Jan. 2017.

175 SZURMANT, H.; WHITE, R. A.; HOCH, J. A. Sensor complexes regulating two-component signal transduction. **Current Opinion in Structural Biology**, v. 17, n. 6, p. 706-715, Dec. 2007.

176 PFAM 28.0. 2015. Disponível em: <<http://pfam.xfam.org/>>. Acesso em: 08 nov. 2015.

177 STEWART, V. The HAMP signal-conversion domain: static two-state or dynamic three-state? **Molecular Microbiology**, v. 91, n. 5, p. 853-857, Mar. 2014.

178 WATTS, K. J. et al. Interactions between the PAS and HAMP domains of the Escherichia coli aerotaxis receptor Aer. **Journal of Bacteriology**, v. 186, n. 21, p. 7440-7449, Nov. 2004.

179 TAYLOR, B. L. Aer on the inside looking out: paradigm for a PAS-HAMP role in sensing oxygen, redox and energy. **Molecular Microbiology**, v. 65, n. 6, p. 1415-1424, Sept. 2007.

180 ZHOU, Q.; AMES, P.; PARKINSON, J. S. Mutational analyses of HAMP helices suggest a dynamic bundle model of input-output signalling in chemoreceptors. **Molecular Microbiology**, v. 73, n. 5, p. 801-814, Sept. 2009.

181 KITANOVIC, S.; AMES, P.; PARKINSON, J. S. Mutational analysis of the control cable that mediates transmembrane signaling in the escherichia coli serine chemoreceptor. **Journal of Bacteriology**, v. 193, n. 19, p. 5062-5072, Oct. 2011.

182 AIROLA, M. V. et al. Structure of Concatenated HAMP domains provides a mechanism for signal transduction. **Structure**, v. 18, n. 4, p. 436-448, Apr. 2010.

183 _____. HAMP domain conformers that propagate opposite signals in bacterial chemoreceptors. **Plos Biology**, v. 11, n. 2, Feb. 2013. doi.org/10.1371/journal.pbio.1001479.

- 184 FERRIS, H. U. et al. The mechanisms of HAMP-mediated signaling in transmembrane receptors. **Structure**, v. 19, n. 3, p. 378-385, Mar. 2011
- 185 PARKINSON, J. S.; HAZELBAUER, G. L.; FALKE, J. J. Signaling and sensory adaptation in Escherichia coli chemoreceptors: 2015 update. **Trends in Microbiology**, v. 23, n. 5, p. 257-266, May 2015.
- 186 ZHOU, Q.; AMES, P.; PARKINSON, J. S. Biphasic control logic of HAMP domain signalling in the Escherichia coli serine chemoreceptor. **Molecular Microbiology**, v. 80, n. 3, p. 596-611, May 2011.
- 187 HULKO, M. et al. The HAMP domain structure implies helix rotation in transmembrane signaling. **Cell**, v. 126, n. 5, p. 929-940, Sept. 2006.
- 188 AMES, P.; ZHOU, Q.; PARKINSON, J. S. HAMP domain structural determinants for signalling and sensory adaptation in Tsr, the Escherichia coli serine chemoreceptor. **Molecular Microbiology**, v. 91, n. 5, p. 875-886, Mar. 2014.
- 189 SCHULTZ, J. E.; KANCHAN, K.; ZIEGLER, M. Intraprotein signal transduction by HAMP domains: a balancing act. **International Journal of Medical Microbiology**, v. 305, n. 2, p. 243-251, Feb. 2015.
- 190 FERRIS, H. U. et al. Mechanism of regulation of receptor histidine kinases. **Structure**, v. 20, n. 1, p. 56-66, Jan. 2012.
- 191 ANANTHARAMAN, V.; BALAJI, S.; ARAVIND, L. The signaling helix: a common functional theme in diverse signaling proteins. **Biology Direct**, v. 1, Sept. 2006. doi: 10.1186/1745-6150-1-25.
- 192 MILLER, A. S.; FALKE, J. J. Side chains at the membrane-water interface modulate the signaling state of a transmembrane receptor. **Biochemistry**, v. 43, n. 7, p. 1763-1770, Feb. 2004.
- 193 KULESEKARA, H. et al. Analysis of Pseudomonas aeruginosa diguanylate cyclases and phosphodiesterases reveals a role for bis-(3'-5')-cyclic-GMP in virulence. **Proceedings of the National Academy of Sciences of the United States of America**, v. 103, n. 16, p. 2839, Apr. 2006.
- 194 GAINES, J. M. et al. Regulation of the Pseudomonas aeruginosa toxA, regA and ptxR genes by the iron-starvation sigma factor PvdS under reduced levels of oxygen. **Microbiology-SGM**, v. 153, p. 4219-4233, Dec. 2007.

195 WALKER, S. L. et al. TOXR (REGA)-mediated in-vitro transcription of pseudomonas-aeruginosa toxa. **Gene**, v. 150, n. 1, p. 87-92, Dec. 1994.

196 HA, D. G.; RICHMAN, M. E.; O'TOOLE, G. A. Deletion mutant library for investigation of functional outputs of cyclic diguanylate metabolism in pseudomonas aeruginosa PA14. **Applied and Environmental Microbiology**, v. 80, n. 11, p. 3384-3393, June 2014.

197 KEARNS, D. B. A field guide to bacterial swarming motility. **Nature Reviews Microbiology**, v. 8, n. 9, p. 634-644, Sept. 2010.

198 AHMAD, I. et al. Complex c-di-GMP signaling networks mediate transition between virulence properties and biofilm formation in salmonella enterica serovar typhimurium. **Plos One**, v. 6, n. 12, Dec. 2011.

199 SAMBROOK, J.; RUSSELL D. W. **Molecular cloning**: a laboratory manual. 3rd ed. New York: Cold Spring Harbor, 2001. v.1-3.

200 NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION. **Distribution of GGDEF, EAL, HD-GYP and PilZ domains in bacterial genomes**. Disponível em: <https://www.ncbi.nlm.nih.gov/Complete_Genomes/c-di-GMP.html>. Acesso em: 27 jul. 2018.

201 BLEICHER, L.; LEMKE, N.; GARRATT, R. C. Using Amino acid correlation and community detection algorithms to identify functional determinants in protein families. **Plos One**, v. 6, n. 12, Dec. 2011. doi: 10.1371/journal.pone.0027786.

202 UNIVERSIDADE FEDERAL DE MINAS GERAIS. Computational Biology. **PFstats**. Disponível em: <<http://www.biocomp.icb.ufmg.br/biocomp/software-and-databases/pfstats/>>. Acesso em: 02 ago. 2018.

203 NEWELL, P. D.; MONDS, R. D.; O'TOOLE, G. A. LapD is a bis-(3',5')-cyclic dimeric GMP-binding protein that regulates surface attachment by Pseudomonas fluorescens Pf0-1. **Proceedings of the National Academy of Sciences of the United States of America**, v. 106, n. 9, p. 3461-3466, Mar. 2009.

204 SIMON FRASER UNIVERSITY. Brinkman Lab. **The Pseudomonas genome database**. 2011. Disponível em: <<http://pseudomonas.com/>>. Acesso em: 08 set. 2018.

205 KANEHISA LABORATORIES. **KEGG**: kyoto encyclopedia of genes and genomes. Disponível em: <<http://www.genome.jp/kegg/>>. Acesso em: 08 nov. 2015

206 FINN, R. D. et al. The Pfam protein families database: towards a more sustainable future. **Nucleic Acids Research**, v. 44, n. D1, p. D279-D285, Jan. 2016.

207 EMBL. **SMART**. 2014. Disponível em: <http://smart.embl-heidelberg.de/smart/set_mode.cgi?NORMAL=1>. Acesso em: 08 nov. 2015.

208 SWISS INSTITUTE OF BIOINFORMATICS. **ProtParam tool**. Disponível em: <<https://web.expasy.org/protparam/>>. Acesso em: 01 ago. 2018.

209 TMHMM Server v. 2.0 prediction of transmembrane helices in proteins. 2015. Disponível em: <<http://www.cbs.dtu.dk/services/TMHMM/>>. Acesso em: 08 set. 2018.

210 PHOBIUS - a combined transmembrane topology and signal peptide predictor. 2015. Disponível em: <<http://phobius.sbc.su.se/>>. Acesso em: 08 set. 2018.

211 NEWBY, Z. E. R. et al. A general protocol for the crystallization of membrane proteins for X-ray structural investigation. **Nature Protocols**, v. 4, n. 5, p. 619-637, 2009.

212 UCL. Department of Computer Science. Bioinformatics Group. **The PSIPRED protein sequence analysis workbench**. Disponível em: <<http://bioinf.cs.ucl.ac.uk/psipred/>>. Acesso em: 08 set. 2018.

213 DROZDETSKIY, A. et al. JPred4: a protein secondary structure prediction server. **Nucleic Acids Research**, v. 43, n. W1, p. W389-W394, July 2015.

214 I-TASSER- protein structure and function prediction. Disponível em: <<http://zhanglab.ccmb.med.umich.edu/I-TASSER/>>. Acesso em: 08 set. 2018.

215 HMMER: biosequence analysis using profile hidden Markov models. 2015. Disponível em: <<http://hmmer.janelia.org/>>. Acesso em: 08 set. 2018.

216 NATIONAL INSTITUTE OF HEALTH. National Center for Biotechnology Information. **Basic local alignment search tool**. 2015. Disponível em: <http://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE=Proteins&PROGRAM=blastp&RUN_PSIBLAST=on>. Acesso em: 08 set. 2018.

217 SKYLIGN.ORG. **SkyLign**: interactive logos for alignments and profile HMMs. 2015. Disponível em: <<http://skylign.org/>>. Acesso em: 08 nov. 2017.

218 KUMAR, S.; STECHER, G.; TAMURA, K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. **Molecular Biology and Evolution**, v. 33, n. 7, p. 1870-1874, July 2016.

219 SCHRÖDINGER. **The PyMOL** molecular graphics system. 2015. Disponível em: <<https://pymol.org/2/>>. Acesso em: 8 nov. 2017.

220 EMSLEY, P. et al. Features and development of Coot. **Acta Crystallographica D**, v. 66, p. 486-501, Apr. 2010.

221 TINA, K. G.; BHADRA, R.; SRINIVASAN, N. PIC: protein interactions calculator. **Nucleic Acids Research**, v. 35, p. W473-W476, July 2007. doi: 10.1093/nar/gkm423.

222 INKSCAPE DRAW FREELY. **Inkscape 0.92**. Disponível em: <<https://inkscape.org/en/>>. Acesso em: 01 ago. 2018.

223 THERMO FISHER SCIENTIFIC. **Oligoperfect primer designer**. 2018. Disponível em: <<https://www.thermofisher.com/br/en/home/life-science/oligonucleotides-primers-probes-genes/custom-dna-oligos/oligo-design-tools/oligoperfect.html>>. Acesso em: 26 set. 2018

224 NEW ENGLAND BIOLABS. **NEBcutter V2.0**. Disponível em: <<http://nc2.neb.com/NEBcutter2/>>. Acesso em: 03ago. 2018.

225 VINCZE, T.; POSFAI, J.; ROBERTS, R. J. NEBcutter: a program to cleave DNA with restriction enzymes. **Nucleic Acids Research**, v. 31, n. 13, p. 3688-3691, July 2003.

226 BUTT, T. R. et al. SUMO fusion technology for difficult-to-express proteins. **Protein Expression and Purification**, v. 43, n. 1, Sept. 2005.

227 AGILENT. **QuikChange Primer Design**. 2018. Disponível em: <<https://www.genomics.agilent.com/primerDesignProgram.jsp>>. Acesso em: 27 jul. 2018.

228 CAMILO, C. M.; POLIKARPOV, I. High-throughput cloning, expression and purification of glycoside hydrolases using Ligation-Independent Cloning (LIC). **Protein Expression and Purification**, v. 99, p. 35-42, July 2014. I

229 LAEMMLI, U. K. Cleavage of structural proteins during assembly of head of bacteriophage-T4. **Nature**, v. 227, n. 5259, p. 680-&, 1970.

230 RYJENKOV, D. A. et al. Cyclic diguanylate is a ubiquitous signaling molecule in bacteria: Insights into biochemistry of the GGDEF protein domain. **Journal of Bacteriology**, v. 187, n. 5, p. 1792-1798, Mar. 2005.

231 MACÊDO, J. N. A. **Estudos estruturais e funcionais de septinas humanas: a ligação e hidrólise de GTP por SEPT3 e a busca de parceiros funcionais de SEPT1, SEPT5 e SEPT7**. 2010. 160 p. Tese (Doutorado em Ciências) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2010.

232 RAO, F. et al. Enzymatic synthesis of c-di-GMP using a thermophilic diguanylate cyclase. **Analytical Biochemistry**, v. 389, n. 2, p. 138-142, June 2009.

233 SHEVCHENKO, A. et al. Mass spectrometric sequencing of proteins from silver stained polyacrylamide gels. **Analytical Chemistry**, v. 68, n. 5, p. 850-858, Mar. 1996.

234 EXPASY PORTAL. **PeptideCutter**. Disponível em: https://web.expasy.org/peptide_cutter. Acesso em: 01 ago. 2018.

235 WEBB, M. R. A continuous spectrophotometric assay for inorganic-phosphate and for measuring phosphate release kinetics in biological-systems. **Proceedings of the National Academy of Sciences of the United States of America**, v. 89, n. 11, p. 4884-4887, June 1992.

236 MOLECULAR PROBES. **EnzChek® Pyrophosphate Assay Kit**. 2003. Disponível em: <https://assets.fishersci.com/TFS-Assets/LSG/manuals/mp06645.pdf>. Acesso em: 28 set. 2018.

237 VOET, D.; VOET, J. G. **Bioquímica**. 3ª ed. Porto Alegre: Artmed, 2006.

238 INMETRO. **Avaliação de dados de medição: guia para a expressão de incerteza de medição**. 2008. Disponível em: http://www.inmetro.gov.br/noticias/conteudo/iso_gum_versao_site.pdf. Acesso em: 27 set. 2018

239 BARTOSCHIK, T. et al. Near-native, site-specific and purification-free protein labeling for quantitative protein interaction analysis by MicroScale Thermophoresis. **Scientific Reports**, v. 8, Mar. 2018. doi.org/10.1038/s41598-018-23154-3.

240 ENTZIAN, C.; SCHUBERT, T. Studying small molecule-aptamer interactions using MicroScale Thermophoresis (MST). **Methods**, v. 97, p. 27-34, Mar. 2016. doi: 10.1016/j.ymeth.2015.08.023.

- 241 SHIBA, K. et al. Polydispersity as a parameter for indicating the thermal stability of proteins by dynamic light scattering. **Analytical Sciences**, v. 26, n. 6, p. 659-663, June 2010.
- 242 JANCARIK, J. et al. Optimum solubility (OS) screening: an efficient method to optimize buffer conditions for homogeneity and crystallization of proteins. **Acta Crystallographica D**, v. 60, p. 1670-1673, Sept. 2004.
- 243 BORGSTAHL; E., G. How to use dynamic light scattering to improve the likelihood of growing macromolecular crystals. In: WALKER, J. M.; DOUBLIÉ, S. **Macromolecular crystallography protocols**. Totowa: Humana Press, 2007. p. 109-129. (Methods molecular biology, v. 363).
- 244 BATTYE, T. G. G. et al. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. **Acta Crystallographica D**, v. 67, Pt. 4, p. 271-281, Apr. 2011.
- 245 KABSCH, W. XDS. **Acta Crystallographica D**, v. 66, p. 125-132, Feb. 2010.
- 246 MARKS, D. S.; HOPF, T. A.; SANDER, C. Protein structure prediction from sequence variation. **Nature Biotechnology**, v. 30, n. 11, p. 1072-1080, Nov. 2012.
- 247 NEWMAN, J. A review of techniques for maximizing diffraction from a protein crystal in stilla. **Acta Crystallographica D**, v. 62, p. 27-31, Jan. 2006.
- 248 ZHANG, Y.; SKOLNICK, J. Scoring function for automated assessment of protein structure template quality. **Proteins-structure function and bioinformatics**, v. 57, n. 4, p. 702-710, Dec. 2004.
- 249 MAO, F. et al. DOOR: a database for prokaryotic operons. **Nucleic Acids Research**, v. 37, p. D459, Jan. 2009. doi: 10.1093/nar/gkn757.
- 250 ZHANG, L. et al. The Escherichia coli CysZ is a pH dependent sulfate transporter that can be inhibited by sulfite. **Biochimica Et Biophysica Acta-biomembranes**, v. 1838, n. 7, p. 1809-1816, July 2014.
- 251 TRALAU, T. et al. Transcriptomic analysis of the sulfate starvation response of Pseudomonas aeruginosa. **Journal of Bacteriology**, v. 189, n. 19, p. 6743-6750, Oct. 2007.
- 252 WHEELER, T. J.; CLEMENTS, J.; FINN, R. D. Skylign: a tool for creating informative, interactive logos representing sequence alignments and profile hidden Markov models. **BMC Bioinformatics**, v. 15, Jan. 2014. doi: 10.1186/1471-2105-15-7.

- 253 SILVA, E. E. D. **GitHub-Tese**. 2018. Disponível em: <<https://github.com/evertonedsilva/GitHub-Tese>>. Acesso em: 27 set. 2018.
- 254 BACHEGA, J. F. R. et al. Systematic structural studies of iron superoxide dismutases from human parasites and a statistical coupling analysis of metal binding specificity. **Proteins: structure function and bioinformatics**, v. 77, n. 1, p. 26-37, Oct. 2009
- 255 FONSECA JÚNIOR, N. J. F. **PFSTATS**: sistema para estudo de famílias de proteínas através de detecção de resíduos conservados e decomposição de redes de coevolução. 2016. 90 p Dissertação (Mestrado em Bioinformática). Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, 2016.
- 256 BOBROV, A. G. et al. Systematic analysis of cyclic di-GMP signalling enzymes and their role in biofilm formation and virulence in *Yersinia pestis*. **Molecular Microbiology**, v. 79, n. 2, p. 533-551, Jan. 2011.
- 257 WYATT, P. J. Light-scattering and the absolute characterization of macromolecules. **Analytica Chimica Acta**, v. 272, n. 1, p. 1-40, Feb. 1993.
- 258 MINTON, A. P. Recent applications of light scattering measurement in the biological and biopharmaceutical sciences. **Analytical Biochemistry**, v. 501, p. 4-22, May 2016.
- 259 GHAZY, R. et al. Static laser light scattering (SLLS) investigations of the scattering parameters of a synthetic polymer. **Optics and Laser Technology**, v. 31, n. 6, p. 447-453, Sept. 1999.
- 260 HASHIM, H. et al. Determination of scattering parameters of polyvinyl alcohol by static laser scattering. **Beni-Suef University Journal of Basic and Applied Sciences**, v. 3, n. 3, p. 203-208, Sept. 2014.
- 261 CORPORATION, W. T. **ASTRA Quick Guide**. Santa Barbara, CA, USA, 2017. Disponível em: <<https://chem.uiowa.edu/sites/chem.uiowa.edu/files/groups/TN1007%20-%20ASTRA%20Quick%20Guide%20Rev%20C.pdf>>. Acesso em: 12/08/18.
- 262 WYATT TECHNOLOGY CORPORATION. **ASTRA quick guide**. 2017. Disponível em: <<https://chem.uiowa.edu/sites/chem.uiowa.edu/files/groups/TN1007%20-%20ASTRA%20Quick%20Guide%20Rev%20C.pdf>>. Acesso em: 12 ago. 2018.
- 263 WIENKEN, C. J. et al. Protein-binding assays in biological liquids using microscale thermophoresis. **Nature Communications**, v. 1, Oct. 2010. doi: 10.1038/ncomms1093.

264 PLATTEN, J. K.; COSTESEQUE, P. Charles Soret. a short biography - on the occasion of the hundredth anniversary of his death. **European Physical Journal E**, v. 15, n. 3, p. 235-239, Nov. 2004.

265 DUHR, S.; BRAUN, D. Thermophoretic depletion follows Boltzmann distribution. **Physical Review Letters**, v. 96, n. 16, p. 168301, Apr. 2006.

266 SEIDEL, S. A. I. et al. Microscale thermophoresis quantifies biomolecular interactions under previously challenging conditions. **Methods**, v. 59, n. 3, p. 301-315, Mar. 2013.

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

268 DIMA, R. I.; THIRUMALAI, D. Determination of network of residues that regulate allostery in protein families using sequence analysis. **Protein Science**, v. 15, n. 2, p. 258-268, Feb. 2006.

269 DEKKER, J. P. et al. A perturbation-based method for calculating explicit likelihood of evolutionary co-variance in multiple sequence alignments. **Bioinformatics**, v. 20, n. 10, p. 1565-1572, July 2004.

APÊNDICES

Apêndice A - Outros resultados do trabalho

Como parte da estratégia de estudo, um grupo inicial de proteínas foi selecionado para que dessas fossem escolhidas aquelas com resultados mais promissores. Duas dessas proteínas são de *Xanthomonas citri* e são codificadas nos genes XAC1348 e XAC1488. Outras proteínas são de *P. aeruginosa* e engloba PA0847, codificada PA14_53310, PA14_37690, correspondente a PA2072 da cepa PAO1.

No total foi planejado trabalhar com 34 construções e com 5 proteínas inteiras, das quais 35 clonagens foram feitas no vetor de expressão pETSUMO, 4 clonagens no vetor expressão para proteínas inteiras pET29, 2 clonagens em um único vetor de coexpressão pETDuet e 6 clonagens no vetor do duplo pTRG. Duas clonagens utilizadas foram providas de um trabalho anterior do nosso laboratório, referentes a PA0847-HPG-1 (pTRG) e HsbD-Nt (pBT). Os primers para essas clonagens estão exibidos no apêndice B. A esquematização da abrangência das sequências originais é mostrada na Figura 89. Os detalhes sobre a abrangência de cada uma das proteínas, assim como os progressos na clonagem, expressão e purificação, estão esquematizados na Tabela 24.

Figura 89 - Esquematização das proteínas e respectivas construções que foram abordadas no trabalho.
Fonte: Elaborada pelo autor.

Tabela 24 - Resumo de resultados de triagens por construções solúveis de diferentes proteínas sensoras transmembranares e de potenciais parceiros de interação.

Proteínas	Abrev.	Domínios	Bases	aa	MM (kDa)	Clon.	Exp.	Purif.
XAC1345	XAC1345	CHASE3+GAF+GGDEF	1710	1-569	62,0			
	C3-1	CHASE3	387	36-164	14,5			
	C3-2	CHASE3	411	36-172	15,5			
	C3-3	CHASE3	432	36-179	16,2			
	GG-1	GAF+GGDEF	1089	207-569	39,2			
	GG-2	GAF+GGDEF	1035	225-569	37,1			
	GG-3	GAF+GGDEF	1017	231-569	36,5			
	GG-4	GAF+GGDEF	945	255-569	33,8			
XAC1488	XAC1488	CHASE3+GAF+GGDEF	1701	1- 566	61,5			
	C3-1	CHASE3	294	35-132	10,9			
	C3-2	CHASE3	408	35-170	15,1			
	C3-3	CHASE3	429	35-177	15,9			
	GG-1	GAF+GGDEF	1080	207-566	38,8			
	GG-2	GAF+GGDEF	1056	215-566	37,8			
	GG-3	GAF+GGDEF	1020	227-566	36,4			
	GG-4	GAF+GGDEF	939	254-566	33,5			
PA2072	PA2072	CHASE4+PAS+GGDEF+EAL	2595	1-864	95,1			
	C4-1	CHASE4	492	46-209	18,0			
	C4-2	CHASE4	525	46-220	19,2			
	C4-3	CHASE4	633	46-256	22,9			
	PGE-1	PAS+GGDEF+EAL	1653	314 -64	61,3			
	PGE-2	GGDEF+EAL	1356	413-864	50,3			
PA0847	PA0847	CHASE4+HAMP+PAS+GGDEF	2046	1-681	76,8			
	C4-1	CHASE4	528	65-240	20,2			
	C4-2	CHASE4	549	65-247	21,0			
	C4-3	CHASE4	711	65-301	27,0			
	C4-4	CHASE4	747	65-313	28,5			
	C4-5	CHASE4	744	65-312	28,4			
	HPG-1	HAMP+PAS+GGDEF	1020	339-678	38,2			
	HPG-2	HAMP+PAS+GGDEF	963	358-678	35,9			
	HPG-3	HAMP+PAS+GGDEF	930	369-678	34,6			
	PG-1	PAS+GGDEF	831	402-678	30,9			
	PG-2	PAS+GGDEF	825	404-678	30,5			
	G	GGDEF	492	517-678	18,3			
	HP-1	HAMP-PAS	492	339-502	18,6			
	HP-2	HAMP-PAS	435	358-502	16,3			
	P	PAS	303	402-502	11,2			
HsbD	HsbD-Nt	N-terminal	153	1-51	6,2			
PA1851	PA1851-Nt	N-terminal	153	1-51	6,3			
RegA	RegA	EAL inativo	780	1-259	28,8			

Legenda: Em azul estão as etapas bem-sucedidas, em amarelo estão etapas cumpridas com alguma dificuldade e em vermelho estão as etapas que precisariam ser otimizadas.

Fonte: Elaborada pelo autor.

Apêndice B - Listas de primers

Nessa seção estão inseridos os primers utilizados para clonagem das construções apresentadas no trabalho, como também de outros experimentos que foram abordados e não se inseriram no trabalho final.

Tabela 25 - Resumo de construções e de primers para produção de proteínas de *X. citri*.

Proteínas	Abrev.	aa	Primers para clonagem em pETSUMO
XAC1345	Inteira	1-569	F: 5'-cagtcataatgatgtcctcttcttttacgtcgcg R: 5'-cagtcctcgaaggccgcattgtccgtcg
	C3-1	36-164	F: 5'-cagtggatcctcgtggtccgatgccg R: 5'-cagtcctcgaagttagcgcgtgcaccatcgcca
	C3-2	36-172	F: 5'-cagtggatcctcgtggtccgatgccg R: 5'-cagtcctcgaagttagcgcgtgtgccagcagctc
	C3-3	36-179	F: 5'-cagtggatcctcgtggtccgatgccg R: 5'-cagtcctcgaagttagcgtttgttcgctgctttgcgc
	GG-1	207-569	F: 5'-cagtggatcccgacgcgaattgcgccacc R: 5'-cagtcctcgaagttagcgcgtgtgtccgtc
	GG-2	225-569	F: 5'-cagtggatcccgagttgggccagagcatcg R: 5'-cagtcctcgaagttagcgcgtgtgtccgtc
	GG-3	231-569	F: 5'-cagtggatcccgacgcgttgccagcgca R: 5'-cagtcctcgaagttagcgcgtgtgtccgtc
	GG-4	255-569	F: 5'-cagtggatccagtgccgaagaagcgttgat R: 5'-cagtcctcgaagttagcgcgtgtgtccgtc
XAC1488	Inteira	1-566	F: 5'-cagtcataatggtgttgaccaaagaaccaaac R: 5'-cagtcctcgaaggccacggatgcaagc
	C3-1	35-132	F: 5'-cagtggatcctcgtgtcggacgccg R: 5'-cagtcctcgaagttagcgttgatagctatgcaataactcgctc
	C3-2	35-170	F: 5'-cagtggatcctcgtgtcggacgccg R: 5'-cagtcctcgaagttagcgggatcagttgctgccg
	C3-3	35-177	F: 5'-cagtggatcctcgtgtcggacgccg R: 5'-cagtcctcgaagttagcgggtgtgtggctgcg
	GG-1	207-566	F: 5'-cagtggatccgagctggaacggcgtagc R: 5'-cagtcctcgaagttagcggccacggatgcaa
	GG-2	215-566	F: 5'-cagtggatccgaagacgacgccagccg R: 5'-cagtcctcgaagttagcggccacggatgcaa
	GG-3	227-566	F: 5'-cagtggatccgagtcgtttccgcactcga R: 5'-cagtcctcgaagttagcggccacggatgcaa
	GG-4	254-566	F: 5'-cagtggatccgatgccgaagaagcgtg R: 5'-cagtcctcgaagttagcggccacggatgcaa

Legenda: preto - sequência complementares aos CDS, desenhadas com o programa OligoPerfect™; azul - sequências dos sítios de restrição; vermelho - códons de parada (*stop codons*). Sítios de restrição: NcoI - ccatgg; NotI - gcggccgc; BamHI - ggatcc; HindIII - aagctt; EcoRI - gaattc; XhoI - ctcgag; NdeI - catatg.

Fonte: Elaborada pelo autor.

Tabela 26 - Resumo de construções e de primers para produção de proteínas de *Pseudomonas aeruginosa* – de proteínas inteiras em pET29 e construções em pETSUMO.

Proteínas	Abrev.	aa	Primers
PA2072 (PA14_37690)	Inteira	1-864	F: 5'-cagtc ccatgg atgcttgcttcagcgattgccagc R: 5'-cagtg cggccgc caggccggcgcgcagcat
	C4-1	46-209	F: 5'-cagtg gatccg acatcgccatcgagcagagcc R: 5'-cagta agctttta gccgagcttggccagcttcg
	C4-2	46-220	F: 5'-cagtg gatccg acatcgccatcgagcagagcc R: 5'-cagta agctttta caaacagggtgcaggttggcgatg
	C4-3	46-256	F: 5'-cagtg gatccg acatcgccatcgagcagagcc R: 5'-cagta agctttta cgcgggtctcccgcagcat
	PGE-1	314-864	F: 5'-cagtg gatccg actggatctgggaagtcgacggg R: 5'-cagtc tcgag tcaaggccggcgcgcga
	PGE-2	413-864	F: 5'-cagtg gatccg acgaggtcgaagcccatgcg R: 5'-cagtc tcgag tcaaggccggcgcgcga
PA0847 (PA14_53310)	Inteira	1-681	F: 5'-cagtc atatg atgtcgaatttcaattcgcac R: 5'-cagtg cggccgc ggctggcgcctggtac
	C4-1	65-240	F: 5'-cagtg gatccg accgagaggaccagcaac R: 5'-cagta agctttta catctgttcgcggagggtt
	C4-2	65-247	F: 5'-cagtg gatccg accgagaggaccagcaac R: 5'-cagta agctttta catcgggttgcagggtcg
	C4-3	65-301	F: 5'-cagtg gatccg accgagaggaccagcaac R: 5'-cagta agctttta gtcgaaacgcagctgcg
	C4-4	65-313	F: 5'-cagtg gatccg accgagaggaccagcaac R: 5'-cagta agctttta ctcgccttggtatgtagagc
	C4-5	65-312	F: 5'-cagtg gatccg accgagaggaccagcaac R: 5'-cagta agctttta cagccctggatgtagagca
	HPG-1	339-678	F: 5'-cagtg gatccg agctgttggttcttcgcg R: 5'-cagta agctttta ctggttacaggcaatagtggttgc
	HPG-2	358-678	F: 5'-cagtg gatccg caacgcccactcgata R: 5'-cagta agctttta ctggttacaggcaatagtggttgc
	HPG-3	369-678	F: 5'-cagtg gatccg gcaacgacgagctggg R: 5'-cagta agctttta ctggttacaggcaatagtggtt
	PG-1	402-678	F: 5'-cagtg gatccg cgacggctacttcgaa R: 5'-cagta agctttta ctggttacaggcaatagtggttgc
	PG-2	404-678	F: 5'-cagtg gatccg gctacttcgaaatggacgtc R: 5'-cagta agctttta ctggttacaggcaatagtggttgc
	G	517-678	F: 5'-cagtg gatccg aaatggcctaccgcgac R: 5'-cagta agctttta ctggttacaggcaatagtggtt
	HP	339-502	F: 5'-cagtg gatccg caacgcccactcg R: 5'-cagta agctttta ctagtcgcggacgatgc
	P	402-502	F: 5'-cagtg gatccg cgacggctacttcga R: 5'-cagta agctttta ctagtcgcggacgatgc
PA3343/HsbD (PA14_20820)	HsbD-Nt	1-51	F: 5'-cagtg gatccg ctgtgctgacacagaaggga R: 5'-cagta agctttta caggccaccacctgatt
PA0707/RegA/ ToxR (PA14-55160)	Inteira	1-259	F: 5'-cagtg aattc atgactgcgacagacagaacgc R: 5'-cagta agctttta cagcaggccggactgct

Legenda: preto - sequência complementares aos CDS, desenhadas com o programa OligoPerfect™; azul - sequências dos sítios de restrição; vermelho - códons de parada (*stop codons*). Sítios de restrição: NcoI - *ccatgg*; NotI - *gcggccgc*; BamHI - *ggatcc*; HindIII - *aagctt*; EcoRI - *gaattc*; XhoI - *ctcgag*; NdeI - *catatg*.
 eeeeeeeFonte: Elaborada pelo autor.

Tabela 27 - Lista de regiões codificantes que foram otimizadas – adquiridos pela empresa GenScript no vetor de propagação pUC57.

PA14-20820/PA3343/HsbD (HsbD-Nt)
Códons otimizados e sítios de restrição
GGATCC
GTGTGTGTTACCCAAAAGACCGCCTGGACGATTCCCGTAGCAACCTGTATGCCGACCAA
CTGAACCGTGGCTTCGTGGCTGCGTTTTCTGCCGGAAGTGAACGTGATTATCGCCGC
TACATGCTGGAAGACTCATTGCCCCTGAAACGCTGA
AAGCTT
Sequência original
GTTTGCGTTACCCAGAAAGATCGTCTGGATGATAGCCGTAGCAATCTGTATGCAGATCAG
CTGAATCGTGGTTTTCTGGTCTGCGTTTTCTGCCGGAAGTGAACGTGATTATCGTCGT
TATATGCTGGAAGATAGCTTTGCACTGAAACGCTGA
PA14-40570/PA1851 região N-terminal – PA1851-Nt
Códons otimizados e sítios de restrição
GGATCC
ATGCTGCCCCGTGATGGTCTGTGCAAGCTGGTCTGCCGGACAACCCGTATGCTCGTCAA
CTGCGTAATGGCTTTTCGTGGCTGCGTTTTGAAAAAGAACTGGAAAATGAATTTCTGTGAA
TTTCTGTCTTGAATAGCCTGATGCAGCGTCGTTGA
AAGCTT
Sequência original
ATGCTGGCACGTGATGGTCTGGTTCAGGCAGGTCTGCCGGATAATCCGTATGCACGTCAG
CTGCGTAATGGTTTTCTGGTGGCTGCGTTTTGAAAAAGAACTGGAAAATGAATTTCTGCGAG
TTTCTGAGCTGGAATAGCCTGATGCAGCGTCGTTGA
Legenda: preto - sequência complementares aos CDS, desenhadas com o programa OligoPerfect™; azul -
sequências dos sítios de restrição; vermelho - trocas dos códons que foram realizadas para otimização da
expressão em <i>E. coli</i> . Sítios de restrição: BamHI - <i>ggatcc</i> ; HindIII - <i>aagctt</i> .
Fonte: Elaborada pelo autor.

Tabela 28 - Sequências dos primers utilizados para realização do protocolo de mutação sítio dirigida no sítio autoinibitório RXXD de PA0847 (R587A).

Primer anti-sense (<i>reverse</i>)	5'-c t t g t c c g g c t g c g c c a g g g t g c t g c g t-3'
Primer sense (<i>forward</i>)	5'-a c g c a g c a c c c t g g c g c a g c c g g a c a a g-3'
Fonte: Elaborada pelo autor.	

Tabela 29 - Resumo de primers para clonagem de construções de PA0847 *P. aeruginosa* para vetores do duplo híbrido utilizando a técnica LIC.

Proteínas	Abrev.	aa	Primers	Planejamento e execução experimental
PA0847	HPG-1	339-678	F: 5'-cagggcgccatggagctgtgggtcctgcg R: 5'-gacccgacgcggtcaggctggcgcc	Andrea Cardoso ¹³
	HPG-3	369-678	F: 5'-cagggcgccatgggcaacgacgagctggg R: 5'-acccgacgcggtcactggtacaggcaatagtgggt	Andrea Cardoso/ Éverton Silva
	PG (PG-1)	402-678	F: 5'-cagggcgccatgcgcgacggctacttca R: 5'-gacccgacgcggtcactggtacaggcaatagtgggt	Andrea Cardoso/ Éverton Silva
	G	517-678	F: 5'-cagggcgccatggaaatggcctaccgcgac R: 5'-gacccgacgcggtcactggtacaggcaatagtgggt	Andrea Cardoso/ Éverton Silva
	HP	339-502	F: 5'-cagggcgccatgcgcaacgcccactcg R: 5'-gacccgacgcggtcactggtacaggcaatagtgggt	Andrea Cardoso/ Éverton Silva
	P	402-502	F: 5'-cagggcgccatgcgcgacggctacttca R: 5'-gacccgacgcggtcactggtacaggcaatagtgggt	Andrea Cardoso/ Éverton Silva
PA14-20820 PA3343 (HsbD)	HsbD-Nt	1-51	F: 5'-cagggcgccatggtgtgcgtgacacagaagg R: 5'-gacccgacgcggtcaggctttcagcgc	Andrea Cardoso ¹³
PA14-40570 PA1851	PA1851-Nt	1-51	F: 5'-cagggcgccatgcttgacgcgacgg R: 5'-gacccgacgcggtcacgcgcgctgc	Andrea Cardoso ¹³
PA14_55160 (RegA)	RegA	1-259	F: 5'-cagggcgccatgactgcgacagacagaacgc R: 5'-gacccgacgcggtcaggcggcgac	Andrea Cardoso ¹³

Legenda: em verde estão as regiões adicionais que permitem a ligação tanto ao vetor pTRG quanto pBT. Em vermelho esta o códon de parada.

Fonte: Elaborada pelo autor.

Tabela 30 - Primers utilizados para sequenciamento dos vetores do duplo híbrido pTRG e pBT.

F: 5'-tgcgccctgagcggcgccggatc
R: 5'-ccgcgtcgggtcaattaatgaactagtgcagctcctctagatcgac

*MCS - *multiple cloning site*.

Fonte: Elaborada pelo autor.

Tabela 31 - Primers utilizados para clonagem em pETDuet das construções de HPG-1 de PA0847 e para construção HsbD-Nt.

Proteínas	Abrev.	aa	Sítio do pETDuet	Primers
PA0847	HPG-1	339-678	MCS-1	F: 5'-cagtggatccggagctgtgggtcctgcgc R: 5'-cagtaagcttcaggctggcgctgtgata
PA14-20820 PA3343 (HsbD)	HsbD-Nt	1-51	MCS2	F: 5'-cagtcatatggtgtgcgtgacacagaagga R: 5'-cagtcctcagtcagcgtttcagcgcaagct

Legenda: preto - sequência complementares aos CDS, desenhadas com o programa OligoPerfect™; azul - sequências dos sítios de restrição; vermelho - códons de parada (*stop codons*); roxo - base adicionada para corrigir o quadro de leitura. Sítios de restrição: NcoI - *ccatgg*; NotI - *gcggccgc*; BamHI - *ggatcc*; HindIII - *aagctt*; EcoRI - *gaattc*; XhoI - *ctcgag*; NdeI - *catatg*.

*MCS - *multiple cloning site*.

Fonte: Elaborada pelo autor.

Apêndice C - Alinhamentos completos de domínios em estudo

Figura 90 - Alinhamento de diversos domínios GGDEF de proteínas já bem estudadas na literatura.

Fonte: Elaborada pelo autor.

Figura 92 - Alinhamento de domínios PAS com responsáveis por captar diferentes estímulos e que são bem caracterizadas.

A numeração mostrada no alinhamento é referente a FixL e *Bradyrhizobium japonicum* (B_jFixL), comumente usada como referência na literatura.

Fonte: Elaborada pelo autor.

Apêndice D - Tabelas de dados obtidos com o PFstats

Tabela 32 - Resultados de correlação (logP) maiores que 70 calculados para o alinhamento de proteínas GGDEF-EAL do banco de dados de Galperin e colaboradores.

Domínio GGDEF						Domínio EAL		
Resíduo 1	Resíduo 2	log(p)	Resíduo 1	Resíduo 2	log(p)	Resíduo 1	Resíduo 2	log(p)
D329(D)	Y439(Y)	80	L302(Y)	R306(Q)	147	I551(I)	Q589(Q)	78
D329(D)	S414(S)	87	R306(Q)	T363(L)	161	C563(C)	Q589(Q)	79
F325(L)	L438(V)	-84	L328(I)	Y439(Y)	71	G555(S)	Q589(Q)	262
F325(L)	A373(V)	-84	D336(D)	Y439(Y)	122	P531(P)	G534(G)	139
F325(L)	M438(V)	112	D336(D)	H340(H)	112	W528(W)	P531(P)	73
S414(S)	Y439(Y)	92	D336(D)	D435(D)	79	H530(D)	P531(P)	166
D327(D)	Y439(Y)	112	G368(G)	Y439(Y)	98	G534(G)	C719(C)	73
D327(D)	H340(H)	91	H340(H)	G368(G)	94	W528(W)	C719(C)	100
D327(D)	K332(K)	87	E371(E)	Y439(Y)	97	G704(G)	C719(C)	71
L302(Y)	A310(L)	76	H340(H)	E371(E)	101	W528(W)	G534(G)	115
A310(L)	T363(L)	71	E371(E)	D435(D)	91	H530(D)	G534(G)	157
R316(L)	R359(R)	133	G369(G)	Y439(Y)	111	G555(S)	C563(C)	78
L302(Y)	R316(L)	75	R359(R)	D362(D)	267	I551(I)	G555(S)	73
R316(L)	T363(L)	81	D362(D)	T363(L)	77	Q565(Q)	W569(W)	133
N335(N)	G343(G)	90	R359(R)	T363(L)	77	L506(L)	W557(W)	85
G445(G)	N447(N)	86	P298(H)	L302(Y)	71	W528(W)	W569(W)	71
D370(E)	K442(K)	70	L302(Y)	T363(L)	158	K659(S)	D682(E)	141
P421(A)	G424(G)	113	H340(H)	Y439(Y)	162	I542(I)	I551(I)	75
K332(K)	Y439(Y)	163	H340(H)	D435(D)	127	I511(I)	E614(E)	-74
K332(K)	D435(D)	90	P421(A)	D423(E)	153	W528(W)	H530(D)	85
K332(K)	H340(H)	120	M438(V)	Y439(Y)	100	L585(I)	F595(L)	-76
K332(K)	D336(D)	110	D435(D)	Y439(Y)	138			
K332(K)	G368(G)	72						

*Os resíduos seguem a numeração da referência TBD1265. O primeiro resíduo é referente ao resíduo correlacionado e o resíduo entre parênteses se refere-se ao resíduo encontrado na sequência de referência.
Fonte: Elaborada pelo autor.

Tabela 33 - Primeira parte dos dados, separados por comunidades, da variação de frequência de resíduos com e sem perturbação.

Comun. 1	% total	Frequências (%) do resíduo da primeira coluna, dada as perturbações abaixo													
		F325(L)	M438(V)	D329(D)	S414(S)	D327(D)	K332(K)	H340(H)	Y439(Y)	L438(V)	D336(D)	G368(G)	D435(D)	E371(E)	G369(G)
F325(L)	35,6	X	56,7	41,7	42,4	42,6	44,7	45,8	49,3	43,9	43,2	43,7	45,4	41,9	42,1
M438(V)	50,8	81	X	58,3	59,8	58,7	62,1	62,4	69,4	60,4	59,7	59,6	59,8	58,1	59
D329(D)	80,3	94	92	X	92,6	89,4	91,2	92,3	95,4	89,6	90,1	89,6	89,8	88,6	88,1
S414(S)	79,2	94,4	93,2	91,4	X	88,6	91,5	92,2	95,5	88,7	90,1	89,4	89,4	87,9	88,1
D327(D)	79,8	95,5	92,1	88,9	89,2	X	93,1	93,7	97,3	89,6	91,1	92	91,9	89,1	88,4
K332(K)	74,9	94,1	91,5	85,1	86,5	87,4	X	91,3	96,9	87,2	89,5	88	89,5	86,1	86,2
H340(H)	73,4	94,5	90,1	84,4	85,4	86,2	89,4	X	95,8	87,7	88,2	88,3	90,5	86,5	85,4
Y439(Y)	61,3	84,9	83,6	72,8	73,9	74,7	79,3	80	X	76	76,3	76,5	79	73,7	74,6
L438(V)	67,2	82,9	79,9	75	75,2	75,5	78,3	80,4	83,3	X	76,9	77,9	78,6	75,3	75,7
D336(D)	77,3	93,9	90,7	86,7	87,8	88,2	92,4	92,9	96,2	88,5	X	89,1	90,7	88,5	87,7
G368(G)	73,1	89,8	85,7	81,6	82,6	84,3	85,9	88	91,3	84,7	84,3	X	86,1	83,7	82,7
D435(D)	72,9	93,1	85,8	81,6	82,3	83,9	87,1	89,9	94	85,3	85,6	85,8	X	85,5	83,8
E371(E)	80,4	94,8	91,9	88,7	89,2	89,8	92,4	94,8	96,7	90,1	92	92	94,3	X	89,7
G369(G)	79,9	94,7	92,8	87,7	89	88,6	92	93,1	97,3	90	90,8	90,4	91,9	89,2	X
Comun. 2	% total	Frequências (%) do resíduo da primeira coluna, dada as perturbações abaixo													
		P298(H)	L302(Y)	R306(Q)	A310(L)	R316(L)	T363(L)	R359(R)	D362(D)						
P298(H)	57,9	X	83,8	71,6	70,3	72,8	76,2	67,7	66,9						
L302(Y)	31,8	45,9	X	60,8	53,6	56,5	67,1	49,9	47						
R306(Q)	39,1	48,4	74,9	X	60,6	62,2	76,7	58,7	56,4						
A310(L)	36,9	44,8	62,2	57,1	X	56,6	61,5	54,8	50,3						
R316(L)	30,6	38,4	54,3	48,5	46,9	X	55,3	56	48						
T363(L)	31,6	41,5	66,6	61,8	52,6	57,1	X	51,1	47,9						
R359(R)	42,5	49,6	66,7	63,7	63,1	77,8	68,8	X	75,1						
D362(D)	52,2	60,3	77,3	75,2	71,2	82,1	79,3	92,4	X						
Comun. 3	% total	Frequências (%) do resíduo da primeira coluna, dada as perturbações abaixo													
		W528(W)	H530(D)	P531(P)	G534(G)	W569(W)	C719(C)	Q565(Q)	G704(G)						
W528(W)	80,8	X	97	94,4	95,9	94,1	95,4	94,2	90,2						
H530(D)	55,8	67	X	77,4	73,7	67,7	67	69,2	62,1						
P531(P)	66,4	77,6	92,2	X	84,2	79,5	78	81,1	72,9						
G534(G)	73,4	87,2	97	93,1	X	87,2	87	88,1	81,5						
W569(W)	66,8	77,8	81,1	80	79,3	X	78,8	93,9	74,3						
C719(C)	71,8	84,7	86,2	84,3	85	84,7	X	86	80,2						
Q565(Q)	43,9	51,2	54,4	53,6	52,7	61,7	52,6	X	48,5						
G704(G)	88,2	98,6	98,2	96,9	98	98,1	98,6	97,6	X						

Fonte: Elaborada pelo autor.

Tabela 34 - Segunda parte dos dados, separados por comunidades, da variação de frequência de resíduos com e sem perturbação.

Comun. 4	% total	Perturbações					Comun. 5	% total	Perturbações		
		I542(I)	I551(I)	G555(S)	Q589(Q)	C563(C)			P421(A)	D423(E)	G424(G)
I542(I)	72,3	X	86,7	83,8	84,9	83,4	P421(A)	55,6	X	88,3	82,4
I551(I)	69,7	83,5	X	85,9	87,2	86	D423(E)	39,1	62,2	X	54,6
G555(S)	60,8	70,5	75	X	91,9	80	G424(G)	42,9	63,6	59,8	X
Q589(Q)	57,9	67,9	72,4	87,5	X	77,2					
C563(C)	53,3	61,4	65,7	70,1	71,1	X					

Comun. 6	% total	Perturb.		Comun. 7	% total	D370(E)	K442(K)	Comun. 8	% total	Perturb.		Comun. 9	% total	Perturb.	
		K659(S)	D682(E)							N335(N)	G343(G)			G445(G)	N447(N)
K659(S)	36,7	X	67,6	D370(E)	85,9	X	94,9	N335(N)	90	X	98,7	G445(G)	74,7	X	95,4
D682(E)	37,3	68,6	X	K442(K)	85,6	94,6	X	G343(G)	87,4	95,8	X	N447(N)	43,4	55,4	X

Comun. 10	% total	Perturb.		Comun. 11	% total	L585(I)	F595(L)	Comun. 12	% total	Perturb.	
		L506(L)	W557(W)							L506(L)	E614(E)
L506(L)	62,3	X	79,8	L585(I)	43,2	X	22,1	I511(I)	30,8	X	11,9
W557(W)	62	79,4	X	F595(L)	40,7	20,8	X	E614(E)	37,2	14,4	X

Fonte: Elaborada pelo autor.

Apêndice E - Clonagem

Uma das estratégias para obtenção de uma proteína de interesse se inicia com a obtenção da região codificante da mesma. Para o isolamento, e em alta quantidade (amplificação), de uma sequência de DNA ou RNA de interesse comumente é utilizada a reação em cadeia da polimerase, abreviada para PCR, do inglês *polymerase chain reaction*. Um dos componentes essenciais da PCR são os primers, oligonucleotídeos de fita simples sintéticos e comerciais. Idealmente, os primers hibridizam de maneira específica, ou seja, se ligam de maneira complementar, em uma região do DNA ou RNA molde. Isso permite que a região de interesse seja sintetizada em inúmeras cópias. No presente estudo, foi utilizada como amostra molde os DNAs genômicos dos microrganismos, disponíveis em nosso laboratório (Figura 93).^z

O DNA de interesse obtido em uma PCR geralmente passa por procedimentos para inserção em um DNA circular chamado plasmídeo modificado ou vetor. Esse processo é chamado clonagem e o vetor final é chamado clone. A composição das sequências de um vetor pode ser extremamente diversa e está relacionada com a aplicação do mesmo e vários deles são comerciais. Quando a função de um vetor é coordenar a bactéria para produção de uma proteína de interesse, ou expressão, termo equivalente, o vetor é chamado de vetor expressão.

Nesse trabalho, a técnica utilizada para inserir o DNA de interesse no vetor de expressão é através de enzimas de restrição. Essas enzimas reconhecem e clivam sequências de DNA palindrômicas e pequenas, geralmente de 6 nucleotídeos, chamadas sítios de restrição. Dois sítios de restrição são usados, um em cada extremidade do DNA de interesse e os mesmos devem estar no vetor, em posições, ordem e região corretas em que se quer fazer a inserção. Na PCR, a polimerização ocorre no sentido 5'-3'. O primer sense, ou *forward*, se liga ao início da região de interesse, na fita 3'-5' do DNA molde produzindo uma fita idêntica a 5'-3' do molde. O primer anti-sense, ou *reverse*, hibridiza com o final da região de interesse e faz o inverso, polimerizando uma cadeia complementar a fita 3'-5' do molde (Figura 93).

^z Técnicas de extração de DNA genômico são muito comuns, inclusive com kits comerciais disponíveis.

Figura 93 - Esquema simplificado dos passos de clonagem utilizando enzimas de restrição.
 Fonte: Elaborada pelo autor.

Apêndice F - Espalhamento de luz acoplado a exclusão molecular (SEC MALS)

Uma das análises através de espalhamento é baseado na variação de intensidade da luz espalhada em função do ângulo de espalhamento ($I(\theta)$). Essas análises são baseadas sobre a intensidade do espalhamento de luz fornecido pela amostra, que sobrepõe o do solvente puro, ou o chamado excesso da taxa Rayleigh ou simplesmente taxa Rayleigh (R_θ). Esse parâmetro é medido através da equação (3)²⁵⁷:

$$R_\theta = f[I_\theta - I_s]/I_0 \quad (3)$$

Essa medida pode ser dada, de maneira alternativa, como expresso em (4):

$$R_\theta = (I_\theta \cdot r^2)/(I_0 \cdot V) \quad (4)$$

Nas equações (3) e (4) cada parâmetro representa:

- f : fator de calibração, que depende do aparato experimental;
- I_θ : Intensidade da luz espalhada de uma amostra medida em um ângulo θ ;
- I_s : Intensidade da luz espalhada no ângulo θ medida com o solvente puro;
- I_0 : Intensidade da luz incidente;
- r : distância entre detector e a amostra;
- V : o volume de espalhamento.

A equação formulada por Zimm que descreve R_θ é derivada da aproximação de Rayleigh-Gans-Debye e é dada por (5):

$$R_\theta = K^*McP(\theta)[1 - 2A_2McP(\theta)] \quad (5)$$

Equação na qual:

- K^* é constante óptica;
- M é a massa molecular do soluto, que pode ser dada em g/mol;
- $P(\theta)$ é função de espalhamento da partícula²⁵⁸, que fornece a dependência angular (θ) do espalhamento de luz;
- A_2 é o segundo coeficiente da expansão virial da pressão osmótica em termos da concentração do soluto, expresso em mol.mL.g⁻². A_2 fornece a dimensão da interação entre soluto e solvente e explica desvios não lineares em concentrações altas do soluto.
- c é a concentração do soluto, representada em g/mL.

Para uma luz incidente polarizada²⁵⁷, a constante K^* é dada por (6):

$$K^* = \frac{4\pi n_0^4}{\lambda_0^4 \cdot N_A} \left(\frac{dn}{dc} \right)^2 \quad (6)$$

Na qual, os termos correspondem a:

- n_0 é o índice de refração do solvente;
- λ_0 é o comprimento de luz utilizado no vácuo;
- N_A é o número de Avogadro, igual a $6,02 \times 10^{23}$;
- $\frac{dn}{dc}$ é o incremento no índice de refração (n) pela concentração de soluto (c), que pode ser fornecido em mL/g. Esse termo é obtido pelo detector de DRI.

Para partículas menores que $\lambda/20^{259-260}$, tem-se que $P(\theta) \approx 1$, ou seja, a molécula espalha luz de maneira isotrópica. Para outros casos, Guiner mostrou que o resultado para $P(\theta)$ calculado para uma molécula de cadeia polimérica aleatória poderia ser simplificado de acordo com (7):

$$P(\theta) = 1 - x, \text{ com } x = \frac{q^2 \langle r_g^2 \rangle}{3} \text{ e } P(\theta) \text{ válida para } x \rightarrow 0 \quad (7)$$

Na equação, q é o vetor de espalhamento e é dado por (8):

$$q = \left(\frac{4\pi n_0}{\lambda_0} \right) \sin \left(\frac{\theta}{2} \right) \quad (8)$$

Com a aproximação obtida na equação (7) é obtido $P(\theta)$ independente do formato ou da conformação da molécula. Essa é uma das vantagens importantes da metodologia. Outro termo apresentado na equação é o raio quadrático médio em relação ao centro de gravidade ($\langle r_g^2 \rangle$), dado pela equação (9):

$$\langle r_g^2 \rangle = \frac{\sum r_i^2 dM}{\sum dM} = \frac{\sum r_i^2 dM}{M} \quad (9)$$

Da equação, $r_g = \sqrt{\langle r_g^2 \rangle}$ é o raio de giro em relação ao centro de massa da partícula. Esse parâmetro fornece uma noção do tamanho da partícula, baseado na distribuição de massa da mesma em torno de seu centro de massa.

Zimm notou que a equação (5) pode ser extrapolada linearmente para uma faixa muito maior de x , dado pela equação (7), e de concentrações do soluto c . Isso é possível reescrevendo

a equação (5) em termos de $P^{-1}(\theta)$ e fazendo uma aproximação para concentrações pequenas, como expresso na equação (10)^{aa,bb}:

$$\frac{K^*c}{R(\theta)} = \frac{1}{MP(\theta)[1 - 2A_2McP(\theta)]} \cong \frac{1}{MP(\theta)} (1 + 2A_2McP(\theta)) \quad (10)$$

Isso fornece a equação de Zimm (11):

$$\frac{K^*c}{R(\theta)} = \frac{1}{MP(\theta)} + 2A_2c \quad (11)$$

Aproximando $P^{-1}(\theta) = \frac{1}{1-x}$ por Taylor na equação (11), com x dado pela equação (5), temos (12):

$$\frac{K^*c}{R(\theta)} = \frac{1}{M} \left(1 + \frac{16\pi^2 \sin^2\left(\frac{\theta}{2}\right) \langle r_g^2 \rangle}{3\lambda_0^2} \dots \right) + 2A_2c \quad (12)$$

O que pode ser simplificado para equação (13):

$$\frac{K^*c}{R(\theta)} = \frac{1}{M} + 2A_2c + \frac{16\pi^2 \sin^2\left(\frac{\theta}{2}\right) \langle r_g^2 \rangle}{3\lambda_0^2 M} \quad (13)$$

Com a plotagem de Zimm (*Zimm plot*), que correlaciona $\frac{K^*c}{R(\theta)}$ com $\sin^2\left(\frac{\theta}{2}\right)$ da equação (13), se obtém o raio de giro da molécula (r_g), o segundo termo virial (A_2) e a massa molecular (M), realizando medidas em diferentes ângulos e diferentes concentrações do soluto. No experimento com a exclusão molecular (SEC-MALS), as concentrações envolvidas geralmente são baixas para ver algum efeito provindo de A_2 .²⁶¹ Portanto, no experimento de SEC-MALS o parâmetro A_2 é configurado como padrão para zero e então o termo $2A_2c$ pode ser desprezado. Dessa forma, a equação (13) pode ser reescrita como exposto em (14):

$$\frac{K^*c}{R(\theta)} = \frac{1}{M} + \frac{16\pi^2 \sin^2\left(\frac{\theta}{2}\right) \langle r_g^2 \rangle}{3\lambda_0^2 M} \quad (14)$$

^{aa} Considerando $2A_2McP(\theta)$ muito menor que 1, é possível expandir em Taylor $\frac{1}{1 - 2A_2M}(\theta)$ e truncar o resultado após o termo linear.

^{bb} Por Taylor, é possível mostrar que $\frac{1}{1-x} = 1 + x + x^2 + \dots$

Figura 94 - Gráfico ilustrando o cálculo de massa molecular e raio de giro médio, através do experimento de SEC-MALS.

De acordo com a equação (14), o termo linear é igual ao coeficiente linear da reta, enquanto do termo angular é possível obter o raio de giro do analito.

Fonte: Elaborada pelo autor.

Outra possível análise em um experimento de espalhamento de luz é baseada nas flutuações da intensidade $I(t)$. Para essa análise, primeiro é necessário apresentar a definição da função de autocorrelação g_2 , função da média das intensidades medidas no tempo t qualquer, $I(t)$, com outra medida a medido um certo tempo depois, definido por τ , com a intensidade definida por $I(t+\tau)$. A variável τ é chamada tempo de atraso (*delay time*), e g_2 é dada como função de τ , dada pela equação (15):

$$g_2(\tau) = \frac{\langle I(t)I(t+\tau) \rangle}{\langle I(t) \rangle^2} \quad (15)$$

O decaimento da função de autocorrelação é exponencial pode ser escrita também pelo movimento de difusão natural das partículas da seguinte equação (16):

$$g_2(\tau) = 1 + \beta e^{-2Dq^2\tau} \quad (16)$$

Nessa equação, β representa o intercepto da função g_2 , que tipicamente é dado pelo limite do equipamento.²⁵⁸ O termo q é referente ao vetor de espalhamento, dado na equação (8), e D representa o coeficiente de difusão translacional, com unidade $\text{cm}^2.\text{s}^{-1}$. Em condições ideais^{262, cc}, pode ser dada pela relação de Stokes-Einstein (17):

$$D = \frac{k_B T}{6\pi\eta r_h} \quad (17)$$

^{cc} O coeficiente de difusão pode ser dado por $D = D_0[1 + kd + K(c^2)]$, equação na qual a constante kd e K fornecem a dependência de D_0 com a concentração do soluto c .

Na equação (17), as variáveis representam:

- k_B é a constante de Boltzmann, com valor de $1,38 \times 10^{-23}$, com unidade de $m^2.kg.s^{-1}.K^{-1}$;
- T representa a temperatura em unidade absoluta em K;
- η é a constante de viscosidade do meio, que pode ser dado em $kg.s^{-1}.m^{-1}$;
- r_h é o raio hidrodinâmico (raio médio da partícula em solução);

Portanto, com a curva de decaimento de g_2 é possível calcular o raio hidrodinâmico em um experimento de espalhamento de luz, dado conhecido as demais constantes biofísicas do solvente e do soluto.

Para demonstrar o decaimento da função g_2 , considere o numerador da divisão dada pela equação (15) Para cada τ , é definida a média de (18):

$$\langle I(t)I(t + \tau) \rangle = \frac{1}{T} \int_0^T I(t)I(t + \tau) dt \quad (18)$$

Escrevendo a intensidade em termos da flutuação sua flutuação $\Delta I(t)$:

$$\langle \Delta I(t) \rangle = 0 \quad (19)$$

$$I(t) = \langle I \rangle + \Delta I(t) \quad (20)$$

É possível reescrever a equação (18) em termos de (19) e (20):

$$\langle I(t)I(t + \tau) \rangle = \langle I \rangle^2 + \langle \Delta I(t)\Delta I(t + \tau) \rangle \quad (21)$$

Para $\tau = 0$, o termo da autocorrelação da equação (21) fica como^{dd}:

$$\langle \Delta I(t)\Delta I(t + \tau) \rangle = \langle \Delta I(t)^2 \rangle = \langle I^2 \rangle - \langle I \rangle^2 \quad (22)$$

Utilizando a equação (22) no resultado de (21), é possível obter (23):

$$\langle I(t)I(t + \tau) \rangle = \langle I \rangle^2, \text{ para } \tau = 0 \quad (23)$$

Para τ suficientemente grande, as variações da intensidade em t e $t+\tau$ perdem a correlação e o termo que é função das variações da intensidade da expressão (21) se anula:

$$\langle \Delta I(t)\Delta I(t + \tau) \rangle = \langle \Delta I(t) \rangle \langle \Delta I(t + \tau) \rangle = 0 \quad (24)$$

Então, a equação (21) assume a seguinte forma, considerando (19):

$$\langle I(t)I(t + \tau) \rangle = \langle I \rangle^2, \text{ para } \tau \rightarrow \infty \quad (25)$$

^{dd} $\langle \Delta I^2 \rangle = \langle (I - \langle I \rangle)^2 \rangle = \langle I^2 - 2I\langle I \rangle + \langle I \rangle^2 \rangle = \langle I^2 \rangle - 2\langle I \rangle \cdot \langle (I) \rangle + \langle I \rangle^2 = \langle I^2 \rangle - \langle I \rangle^2$

Figura 95 - Ilustração do processamento de dados para o cálculo do raio hidrodinâmico (r_h).

A função de autocorrelação é obtida calculando as médias das intensidades em t e $t + \tau$; B) Se as intensidades perdem correlação a média do produto de $\Delta I(t)$ e $\Delta I(t+\tau)$ pode ser reescrita como produto das médias desses termos, que são nulas. Assim, resta apenas o termo da média ao quadrado de $I(t)$; C) Dividindo o termo da ordenada plotada em (B) por $\langle I(t) \rangle^2$ se obtém g_2 . É possível relacionar a função de autocorrelação em termos da difusão das partículas; D) É possível obter o raio de hidrodinâmico da partícula através do termo de difusão da função g_2 .

Fonte: Elaborada pelo autor.

Apêndice G - Parâmetros de cinética e inibição enzimáticas

A equação de Michaelis-Menten descreve a atividade de uma enzima no estado quase-estacionário (26). A equação descreve a velocidade inicial (V_0) de conversão do substrato por uma enzima em função da concentração do substrato $[S]$ e de dois parâmetros característicos de cada enzima, da velocidade máxima de conversão ($V_{\text{máx}}$) e da constante de Michaelis-Menten (K_m). O K_m corresponde a concentração de substrato que fornece uma velocidade igual a metade da velocidade máxima da enzima. É um parâmetro que é diretamente proporcional a afinidade da enzima pelo substrato.

$$V_0 = \frac{V_{\text{máx}} \cdot [S]}{K_m + [S]} \quad (\mu\text{M} \cdot \text{s}^{-1}) \quad (26)$$

As concentrações de substrato utilizadas para cálculos devem estar idealmente distribuídas em torno do valor de K_m . Isso se faz necessário para que se obtenha bons parâmetros para cálculo da curva (hipérbole quadrática). Para obter o K_m pode-se recorrer a valores já descritos na literatura de enzimas homólogas a que esta em estudo ou através de testes preliminares. Os passos para o cálculo estão ilustrados a seguir (Figura 96).

Com o valor de $V_{\text{máx}}$ é possível calcular a constante catalítica k_{cat} , que fornece a eficiência enzimática, ou seja, número de moléculas de substrato consumido normalizado por unidade de enzima (27).

$$K_{\text{cat}} = \frac{V_{\text{máx}}}{[\text{Enzima}]} \quad (\text{s}^{-1}) \quad (27)$$

Já para quantificação da inibição da atividade enzimática por uma molécula I é através de uma curva de dose e resposta. Com valores da velocidade inicial de conversão em função da concentração do inibidor ($V_0(I)$), são calculadas as taxas de redução $V_0\%(I)$. Essas velocidades são em relação a uma reação controle, sem adição do inibidor $V_0(0)$, como exposto a seguir (28):

$$V_0\%(I) = 100 \left(1 - \frac{V_0(I)}{V_0(0)} \right) \quad (28)$$

Com os valores obtidos com da expressão (28) é graficado a função de quatro parâmetros ou a função de dose e resposta (29):

$$V_0\%([I]) = A1 + \frac{A2 - A1}{1 + 10^{(\log IC_{50} - [I])h}} \quad (29)$$

Na equação A1 e A2 são as assíntotas horizontais inferior e superior respectivamente. A assíntota inferior deve ser, idealmente, próxima de 0 %, enquanto a assíntota superior deve estar por volta de 100 %, indicando a saturação do inibidor e consequentemente a eliminação da atividade com a adição do mesmo. O coeficiente de Hill é representado por h. De maneira análoga ao K_m , as concentrações do inibidor idealmente devem ser igualmente distribuídas em torno do valor da concentração definida pelo IC_{50} .

Figura 96 - Esquematização dos procedimentos para cálculo de K_m e $V_{\text{máx}}$.

Previamente é necessário determinar uma concentração mínima de enzima suficiente para ser observado sinal acima de ruídos experimentais. Nos ensaios, meios reacionais com diferentes concentrações de substrato são monitorados no tempo. Os dados de formação do produto podem ser de maneira direta ou indireta de acordo com a técnica que consiga detectar apenas o produto ou apenas o substrato. A mais comum é monitorar a absorbância e usar uma curva padrão para convertê-la em concentração de substrato. As regiões lineares iniciais das curvas das reações são usadas para cálculo da taxa de formação de produto pelo tempo, velocidade inicial (V_0). Essa região inicial é válida a condição de estado estacionário, umas das condições do modelo da cinética de Michaelis-Menten. Com os dados de velocidade é confeccionado o gráfico de Michaelis-Menten, que fornece os parâmetros K_m e $V_{\text{máx}}$.

Fonte: Elaborada pelo autor.

Apêndice H - Termoforese

O fenômeno do movimento de partículas sobre um gradiente de temperatura, chamado de efeito Soret, ou termoforese, teve suas primeiras descrições feitas por Carl Ludwig, em 1856²⁶³, e Charles Soret, em 1879.²⁶⁴ O fluxo de partículas/solutos promovido por um gradiente de temperatura pode ser dado por (30)²⁶⁵:

$$j_T = -cD_T \nabla T \quad (30)$$

Equação na qual:

- c denota a concentração do soluto;
- ∇T indica o gradiente de temperatura;
- D_T é referente ao termo de difusão frente a um gradiente de temperatura;

Por outro lado, o fluxo de partículas também é regido pela difusão natural da mesma, que depende da concentração da mesma (31).

$$j_D = -D \nabla c \quad (31)$$

Em um sistema estacionário, i.e., em equilíbrio termodinâmico, esses fluxos se compensam e se anulam, fornecendo a equação a seguir (32):

$$j = j_D + j_T = -D \nabla c - cD_T \nabla T = 0 \rightarrow D \nabla c = -cD_T \nabla T \quad (32)$$

Rearranjando a equação (32) em termos dos infinitesimais, obtém-se (33):

$$\frac{dc}{c} = -\frac{D_T}{D} dT \quad (33)$$

Da equação (33) se define o coeficiente de Soret (34), que é referente à razão entre os termos de difusão:

$$S_T = -\frac{D_T}{D} \quad (34)$$

Integrando a equação e substituindo o coeficiente de Soret, obtém-se a equação (35)^{ee}:

$$\ln \left(\frac{c}{c_0} \right) = -S_T (T - T_0) \rightarrow \frac{c}{c_0} = e^{-S_T (T - T_0)} \quad (35)$$

Portanto, em um sistema no qual ocorre um gradiente de temperatura é observada uma concentração de soluto que cai exponencialmente no sentido do gradiente. A termoforese possui uma teoria que ainda está em debate. Como consenso, é postulado que S_T é função do tamanho da partícula e de parâmetros da superfície da mesma, como carga e hidrofobicidade.²⁶³

^{ee} $\int_i^f \frac{dx}{x} = \ln x_f - \ln x_i = \ln \frac{x_f}{x_i}$

Em um experimento com partículas fluorescentes, como a abordada no presente texto, a fluorescência é proporcional à concentração das mesmas (36).²⁶⁶

$$F_{norm} = \frac{F}{F_0} = \frac{c}{c_0} \quad (36)$$

Na equação (36), F_0 representa a fluorescência medida sem o gradiente de temperatura ou na temperatura sem o aquecimento (T_0), referenciada como temperatura antes do *T-jump*.²⁶⁶ Já a variável F é a fluorescência da partícula na maior temperatura T . F_{norm} representa a fração da fluorescência final em relação a inicial ou, de maneira equivalente, a razão entre as concentrações sem e com mudança de temperatura. A variação da fluorescência é proporcional à sua concentração em termos das frações com e sem a diferença de temperatura, e pode ser dada por ΔF_{norm} .(37):

$$\Delta F_{norm} = 1 - F_{norm} \quad (37)$$

É possível comparar a equação (37) com aquela obtida em (35). Para isso, faz-se uma aproximação por expansão^{ff} da equação (37), para variações de temperaturas pequenas, tem-se que:

$$\frac{F}{F_0} \approx 1 - S_T(T - T_0) \quad (38)$$

Dessa maneira é possível correlacionar a queda de fluorescência com o coeficiente de Soret. Vale lembrar que é necessário acrescentar nos cálculos as mudanças da fluorescência no fluoróforo devido a mudança de temperatura $(\delta F / \delta T)$, se houver:

$$\frac{F}{F_0} \approx 1 - S_T(T - T_0) - \frac{\delta F}{\delta T}(T - T_0) = 1 - \left[\left(S_T + \frac{\delta F}{\delta T} \right) (T - T_0) \right] \quad (39)$$

Portanto, é possível comparar as equações (38) e (39) para mostrar que a variação na taxa de fluorescência normalizada está diretamente relacionada com o coeficiente de Soret:

$$\Delta F_{norm} \approx \left(S_T + \frac{\delta F}{\delta T} \right) (T - T_0) \quad (40)$$

^{ff} A expansão em torno de $x=0$ da função é da forma $e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots$

Apêndice I - A termoforese em microescala no cálculo da constante de dissociação (K_d)

Um resultado importante que pode ser obtido através do MST é sobre a quantificação da interação entre dois parceiros moleculares. No experimento, um dos parceiros de interação é marcado com fluoróforo e utilizado em uma concentração fixa $[M]$. Então são realizadas diferentes medidas com diferentes concentrações do ligante $[L]$. Dado o equilíbrio de (41):

A constante de dissociação (K_d) é definida por (42):

$$K_d = \frac{[L_{livre}][M_{livre}]}{[ML]} \quad (42)$$

No experimento é possível realizar uma regressão da fração ligada, F_b , da molécula. Isso porque em medidas com concentrações baixas do ligante L , o sinal observado é provindo majoritariamente da molécula marcada M . Em concentrações elevadas do ligante, o equilíbrio é deslocado para forma complexada e o sinal então é provindo quase que exclusivamente pelo complexo entre molécula marcada e ligante (ML) (Figura 97).

$$F_b = \frac{[ML]}{[M]} \quad (43)$$

Portanto, para reescrever a fração ligada F_b (*fraction bound*) em termos das concentrações do experimento, $[L]$ e $[M]$, é necessário reescrever $[ML]$ em termos dessas concentrações e da constante K_d , que se deseja obter. Sabendo que as concentrações totais do ligante ($[L]$) e da molécula marcada testada ($[M]$), podemos obter que:

$$[L] = [L_{livre}] + [ML] \rightarrow [L_{livre}] = [L] - [ML] \quad (44)$$

$$[M] = [M_{livre}] + [ML] \rightarrow [M_{livre}] = [M] - [ML] \quad (45)$$

Pode-se substituir as equações (44) e (45) na equação (43) para obter:

$$K_d = \frac{([L] - [ML])([M] - [ML])}{[ML]} = \frac{([M][L] - [ML][L] - [ML][M] + ([ML])^2)}{[ML]} \quad (46)$$

Rearranjando a equação (46), se obtém:

$$[ML]^2 - [ML]([L] + [M] + K_d) + [M][L] = 0 \quad (47)$$

Portanto, resolvendo a equação de segundo grau de (47), tem-se:

$$[ML] = \frac{[L] + [M] + K_d \pm \sqrt{([L] + [M] + K_d)^2 - 4[M][L]}}{2} \quad (48)$$

Combinando o resultado de (48) com a fração ligada expressa em (43), obtemos uma expressão que relaciona a curva experimental com o valor de K_d :

$$F_b = \frac{[ML]}{[M]} = \frac{[L] + [M] + K_d - \sqrt{([L] + [M] + K_d)^2 - 4[M][L]}}{2[M]} \quad (49)$$

Para reações em que mais de 1 ligante se associa por molécula marcada ou em caso de ligação cooperativa é utilizada a equação de Hill, escrita como:

$$F_b = \frac{[ML]}{[M]} = \frac{1}{1 + \frac{K_d}{[L]^h}} \quad (50)$$

Na equação (50), h é o coeficiente de Hill.²⁶³ Esse modelo é válido somente quando a concentração ligada é desprezível, ou seja, $[L_{\text{livre}}] \approx [L]$.

Figura 97 - Exemplificação ilustrativa das curvas e dos cálculos do experimento de MST para cálculo de K_d .
Fonte: Elaborada pelo autor.

Apêndice J - Análise de coevolução de resíduos

A análise utilizada nesse texto é fundamentada em uma pontuação, dada através de uma das métricas de correlação, baseadas em abordagens por perturbação. A métrica aplicada, assim como o método utilizado, foi proposto por Bleicher e colaboradores, inspirados em trabalhos de Lockless e Ranganathan²⁶⁷, Dima e Thirumalai²⁶⁸ e Dekker e colaboradores.²⁶⁹ A perturbação na análise é referente ao efeito da presença de um aminoácido em uma posição do alinhamento sobre a distribuição de aminoácidos em outra posição.²⁰¹ A métrica apresentada por Lockless e adaptada por Dima foram denominadas de acoplamento estatístico (SCA, abreviação *statistical coupling analysis*) com uma abordagem análoga a da energia livre de um sistema. Já a métrica de Dekker utiliza uma estatística convencional com cálculos de covariância.

Antes dos cálculos é executado, de maneira opcional, o procedimento a filtragem de sequências, na qual podem ser excluídas aquelas com identidade acima de um valor a , abaixo de um outro valor b ou com uma cobertura em relação a uma sequência de referência menor que c , sendo que os valores de corte a , b e c , além da sequência de referência, são fornecidos pelo usuário. A segunda etapa consiste na escolha da fração mínima de sub-alinhamento em relação ao alinhamento total, de maneira que o sub-alinhamento ainda represente o alinhamento original. Essa escolha é baseada no cálculo da entropia da informação, chamada também de entropia de Shannon, para diferentes sub-alinhamentos. Para cada sub-alinhamento é calculada a entropia de cada posição i , de acordo com equação (51).

$$S_i = - \sum_{x=1}^{20} p_x \log_2(p_x) \quad (51)$$

Sendo que a probabilidade p_x é a probabilidade de cada posição i possuir o aminoácido x é dada por (52):

$$p_x = \frac{n_x}{N} \quad (52)$$

Então é realizado o cálculo da média da entropia de todas as posições N do sub-alinhamento (S_{sa}), como descrito em (54):

$$\langle S_{sa} \rangle = \frac{\sum_{i=1}^N S_i}{N} \quad (53)$$

A entropia de sub-alinhamentos cada vez menores é aproximadamente a mesma, até que, dado certo tamanho da amostra original de sequências, é observado uma mudança o valor

da entropia considerável, o que indica a não representatividade a partir daquele tamanho de sub-alinhamento.

Determinado o valor mínimo é realizado o procedimento do cálculo de correlação. O alinhamento filtrado, com N sequências, tem as posições analisadas aos pares, designadas genericamente como i e j . Em cada uma das posições, cada um dos 20 resíduos de aminoácidos são avaliados. Generalizando uma das rodadas do programa, são encontrados n_A sequências com o resíduo X na posição i e n_B sequências com o resíduo Y na posição j . Considerando que, no alinhamento total, o resíduo Y aparece com uma taxa igual a $\frac{n_B}{N}$. Em um sub-alinhamento contendo somente as sequências com X na posição i , portanto com número n_A de sequências, a frequência do resíduo Y deve ser diferente de $\frac{n_B}{N}$, caso houver uma relação entre os resíduos X e Y nas posições i e j (exemplo na Figura 98A).

Figura 98 - Exemplificação esquemática das análises feitas com o PFstats usando um alinhamento modelo.

A) O programa analisa por vez todas as combinações de posições possíveis. Dado a presença do aminoácido X na posição i do alinhamento, o programa contabiliza qual a frequência do aminoácido Y na posição j . Caso a frequência do resíduo Y nesse novo alinhamento seja maior que um corte mínimo, fornecido pelo usuário, essas posições seguem para análises de pontuação ($\log(P)$). No exemplo apresentado, tem-se $X=R$ (arginina), na posição 147 e $Y=D$ (aspartato), na posição 169. No exemplo, 75 % das sequências com R147 também possuem D169; B) Dado as posições relevantes de acordo com as pontuações, segue-se para análise funcional dos resíduos. É feita uma correspondência entre a posição do alinhamento com as sequências de referência, para elucidar as razões para acorrer a coevolução dos resíduos. No exemplo mostrado, os resíduos acoplados correspondem a R359 e D169, do sítio inibitório RXXD de GGDEFs.

Fonte: Elaborada pelo autor.

No método, o usuário precisa fornecer, além da fração mínima do alinhamento, a pontuação mínima ($\log(p)$; equações (55) e (56)) para aceitar uma correlação como válida e a frequência mínima que o resíduo Y na posição j deve ter no sub-alinhamento com X na posição i . Por exemplo, se for escolhido 0,3, a frequência do resíduo deve ser maior que 70 % na restrição com o resíduo X em i . No caso de anti-correlação, o resíduo Y deve possuir uma taxa menor que 30 % no exemplo dado. Respeitada a condição da frequência mínima, o programa procede com o cálculo da métrica de correlação p . Ele é dado pela cumulativa da distribuição binomial (54):

$$p = \sum_{n=n_i}^{n_f} \frac{n_A}{n! (n_A - n)!} \left(\frac{n_B}{N}\right)^n \left(1 - \frac{n_B}{N}\right)^{n_A - n} \quad (54)$$

No caso em que a frequência $\frac{n_B}{N}$ no subalinhamento com n_A sequências ultrapassa certo valor, há o indício de uma correlação positiva. Então, o número de eventos com sucesso (de se obter Y) são contabilizados de $n_i = n_{B|A}$ até $n_f = n_A$, dado n_A eventos totais (tentativas). No caso de eventos com correlação negativa, ou anti-correlação, os limítrofes da somatória, n_i e n_f , são 0 e $n_{B|A}$, respectivamente. A pontuação é dada pelas equações (55) e (56):

$$-\log(p), \text{ no caso de uma correlação positiva} \quad (55)$$

$$\log(p), \text{ no caso de uma correlação negativa} \quad (56)$$

Os pares de resíduos com valor de correlação foi maior que o estipulado pelo usuário, são agrupados caso compartilhem resíduos. Esses grupos são denominados comunidades e podem ser representadas por redes, nas quais os nós, ou vértices, representam os resíduos e as arestas, ou conexões, representam as correlações (exemplo da Figura 98B). Uma possível interpretação estrutural pode ser feita a partir dos dados de correlação entre resíduos (exemplo da Figura 98B).

O trabalho original de Bleicher²⁰¹ propõe ainda cálculo da correlação entre as comunidades e quantifica a afinidade de uma sequência por cada uma das comunidades. Dada as comunidades A e B , com números de nós com iguais a N_A e M_B e o conjuntos de nós (resíduos) dados por $A = [a_1, a_2, \dots, a_n]$ e $B = [b_1, b_2, \dots, b_m]$. A quantidade Δ_{AB} , que mede a relação entre as comunidades A e B , é dada por (57):

$$\Delta_{AB} = \frac{1}{N_A M_B} \sum_{a_n, b_m} w(a_n, b_m) \quad (57)$$

Na qual $w(a_n, b_n)$ denota o módulo da correlação entre o resíduo a_n da comunidade A e b_n da comunidade B.

Para quantificar quão próximo uma sequência específica S se enquadra em uma dada comunidade A, com número de resíduos N_A , é utilizado o parâmetro aderência (Adh). Esse parâmetro é calculado como:

$$\text{Adh}(S, A) = \frac{1}{N_A(N_A - 1)} \sum_{a_i, a_j \in A} w(a_i, a_j) \delta(a_i, a_j) \quad (58)$$

Nessa equação na qual $\delta(a_i, a_j)$ é igual a 1, quando os resíduos da comunidade a_i e a_j estão presentes na sequência S e igual a 0, quando qualquer um desses elementos não estão presentes. O termo $w(a_i, a_j)$ é o valor da correlação entre a_i e a_j .