UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE QUÍMICA

Programa de Pós-Graduação em Ciências Biológicas (Bioquímica)

VALQUIRIA PIANHERI SOUZA

Papel das Redes Estruturais Proteicas nas Propriedades de
uma Beta-Glicosidase

Versão original da Tese defendida

São Paulo

Data de Depósito na SPG:
05/07/2017

VALQUIRIA PIANHERI SOUZA

Papel das Redes Estruturais Proteicas nas Propriedades de
uma Beta-Glicosidase

Tese apresentada ao Instituto de Química da
Universidade de São Paulo para obtenção do
Título de Doutor em Ciências (Bioquímica)

Orientador: Prof. Dr. Sandro Roberto Marana

São Paulo
2017

Ficha Catalográfica
Elaborada pela Divisão de Biblioteca e
Documentação do Conjunto das Químicas da USP

S719p Souza, Valquiria Pianheri
Papel das redes estruturais proteicas nas
propriedades de uma beta-glicosidase / Valquiria
Pianheri Souza. - São Paulo, 2017.
303 p.

Tese (doutorado) - Instituto de Química da
Universidade de São Paulo. Departamento de
Bioquímica.
Orientador: Marana, Sandro Roberto

1. redes de estrutura proteica. 2. enzimas. 3.
estabilidade térmica. 4. beta-glicosidases. 5.
espalhamento de raios-X a baixos ângulos. I. T. II.
Marana, Sandro Roberto, orientador.

Dedico esta tese aos meus pais, Claudio e Rosana, e a meu noivo, Armando, que com muito entusiasmo e amor, me motivaram a alcançar horizontes que nunca teria imaginado.

AGRADECIMENTOS

Primeiramente a Deus, por me oferecer oportunidades que me trouxeram até aqui.

Ao meu orientador, prof. Sandro Roberto Marana, meu pai científico, por me presentear com este motivante tema de trabalho, e por todos os ensinamentos e discussões fundamentais para a minha formação. A todos os membros do grupo de Enzimologia e agregados, do passado e do presente, por terem tornado esse percurso alegre e com momentos únicos, que nunca vou esquecer. Em especial, agradeço a Maira Frutuoso, Vitor Medeiros, Daniela Beton, Lúcio Mendonça, Éverton Menezes, Fábio Tamaki, Larissa Gonçalves, Bruno, André Reis, Mateus Galdino, Fernanda Matias, Roberto Rozenberg e Raphael Sayegh, por toda amizade, discussões e apoio, muito importantes para a realização desta tese. A Cecília Ikegami, que iniciou este trabalho, pelo apoio científico e também pela amizade. A Thabata Vinha, a primeira estudante de Iniciação Científica que ajudei a orientar, que certamente terá grande sucesso em sua carreira.

A amizade e suporte técnico oferecidos por Edson, José Galeote e Juliana Raveli.

Ao prof. Guilherme Menegon, cuja produção da rede da Sfβgli foi fundamental para o desenvolvimento desta tese, e também aos seus ensinamentos na disciplina de Biologia Estrutural. Ao prof. Chuck Farah, também por seus ensinamentos na disciplina de Biologia Estrutural, e pelo apoio dado ao grupo na cristalografia da Sfβgli.

Ao prof. Cristiano Oliveira, do IFUSP, por todos os ensinamentos de “Espalhamento de Raios-X a baixos ângulos – SAXS” e por todo apoio e discussões esboçadas nesta tese. A todos os demais membros do grupo de fluidos complexos, pelo apoio e amizade, em especial a Renata Bicev e Dennys Reis, que auxiliaram na coleta dos dados apresentados nesta tese.

A Central Analítica, por viabilizar o uso do Espectropolarímetro (CD) e ao Marcio Nardelli, por todas as dicas e suporte para coleta dos dados de CD. Ao Dr. Diorge Pereira, por me auxiliar com sua experiência na coleta de dados de CD de proteínas.

A prof. Carla Columbano, e ao seu grupo, em especial ao Dr. Rogério Lourenço, Dra. Marcela Bath e Maria Griselda, pelo acesso e apoio no uso do PCR real time, e da análise de ThermoFluor.

A todos os integrantes do laboratório de Bioquímica de Insetos, em especial a Cristiane, Renata, Gilliard, Ignácio, Larissa, André e Otto, pela recepção, amizade e por toda a ajuda oferecida durante estes anos de doutorado. Aos profs. Clélia Ferreira e Walter Terra, pelos ensinamentos de Enzimologia apresentados na disciplina de Cinética Enzimática, além das sugestões que contribuíram muito com este trabalho. A Ticiane Damasceno e Daniela Bataglioli, minhas amigas desde a época de graduação, por toda a amizade e apoio oferecidos durante esta etapa.

A prof. Iolanda Cuccovia, por abrir o seu laboratório, permitindo que eu utilizasse diversos equipamentos e reagentes, importantes para a construção desta tese. A Márcia, Katia e Phelipe pelo apoio oferecido. A prof. Maria Teresa Macchini, pela amizade e por todo apoio oferecido desde a época de graduação. A prof. Silvia Agostinho, pelos ensinamentos na disciplina de Termodinâmica, e também pela amizade. Ao prof. Bayardo Torres, por despertar em mim um grande interesse em educação, que me levaram a cursar licenciatura.

A FAPESP, ao CNPq e ao Insituto de Química, pelas bolsas e suportes financeiros.

A toda a minha família, em especial aos meus pais, Rosana e Claudio, e ao meu noivo Armando, peças fundamentais que contribuíram muito em minha carreira e que me apoiaram sempre em todas as decisões. A todas as pessoas que não foram aqui citadas, mas fizeram parte desta caminhada.

Muito obrigada a todos!

*“Onde há vida, existem padrões, e onde há um padrão, existe matemática. É impossível
haver um Universo não matemático contendo observadores vivos.”*

(John D. Barrow, traduzido de “The Artful Universe”).”

RESUMO

Souza, V. P.; Papel das Redes Estruturais Proteicas nas Propriedades de uma Beta-Glicosidase. 2017. 303p. Tese - Programa de Pós-Graduação em Bioquímica. Instituto de Química, Universidade de São Paulo, São Paulo.

A análise de proteínas como redes é uma ferramenta poderosa para compreender as suas propriedades e a importância relativa de seus resíduos. Nesta análise, os resíduos que interagem entre si, covalentemente ou não, são chamados conectados. Nesta abordagem, alguns resíduos contribuem mais fortemente para manter as propriedades da rede, sendo chamados de centrais. Diversos trabalhos têm apontado que resíduos centrais da Rede de Estrutura Proteica também são importantes nas propriedades das proteínas, desempenhando papéis na catálise, estabilidade térmica e alosteria. No entanto, existe falta de trabalhos desenhados de forma sistemática para confirmar esta hipótese. Neste sentido, esta tese tem como objetivo avaliar se existe correlação entre a centralidade dos resíduos de uma enzima, a beta-glicosidase de *Spodoptera frugiperda*, Sfβgli, e a importância destes resíduos na determinação das suas propriedades.

Para isso, foram utilizadas duas abordagens (capítulo 1): Na primeira, os resíduos centrais foram diretamente perturbados substituindo-os, através de mutação sítio-dirigida, por alanina. Na segunda, perturbações no resíduo central foram feitas modificando a vizinhança deste resíduo através de mutações que introduziram ou removeram volume de seu entorno. A partir disso, foi avaliado se estas perturbações afetaram as propriedades Sfβgli.

De forma geral, foi observado (capítulo 2) que as perturbações nos resíduos centrais por ambas abordagens afetam significativamente a termoestabilidade da proteína, reduzindo a sua T_m em até 15°C e aumentando a velocidade de sua desnaturação térmica em até mais de 20 vezes. Além disso, a atividade catalítica de Sfβgli é reduzida por estas

perturbações (capítulo 3), sendo que este efeito e a perda da termoestabilidade parecem resultar da mesma causa, a perturbação do resíduo central.

No capítulo 4, a investigação do estado oligomérico da Sfβgli por SAXS revelou que esta ocorre preponderantemente como dímero em citrato-fosfato 100 mM pH 6,0, mas como um grande oligômero, possivelmente um dodecâmero, em fosfato 10 mM pH 6,0. Paralelamente foi demonstrado que Sfβgli passa por uma ativação quando em tampão fosfato 10 mM, convergindo para as propriedades cinéticas de Sfβgli em citrato-fosfato 100 mM. Redes de Estrutura Proteica foram produzidas considerando-se também a interação entre as cadeias polipeptídicas constituintes de oligômeros de Sfβgli (dímeros, tetrâmeros e hexâmeros). Assim, observou-se que cinco resíduos são sempre centrais por betweenness, mesmo considerando diferentes oligômeros da Sfβgli. Destes, E187, P188 e N329 desempenham papéis conhecidos na catálise e S247 e N249 foram caracterizados nesta tese.

Por fim, no capítulo 5, analisando a centralidade dos resíduos da Rede Estrutural da Sfβgli, observa-se uma preponderante presença de resíduos centrais por $C_{\Delta Lp}$, closeness e betweenness no topo do beta-barril, demonstrando que esta região é muito próxima dos demais resíduos da proteína. Além disso, uma análise da centralidade dos resíduos de 21 beta-glicosidases GH1 revelou que resíduos centrais por closeness são bastante conservados, sendo encontrados predominantemente no sítio ativo destas enzimas, enquanto que dentre os centrais por betweenness há variabilidade.

Portanto os resultados apresentados nesta tese suportam experimentalmente a hipótese de que a centralidade dos resíduos na Rede de Estrutura Proteica é correlacionada com propriedades funcionais das proteínas.

Palavras-chave: Redes de estrutura proteica; enzimas; estabilidade térmica; beta-glicosidases; espalhamento de raios-X a baixos-ângulos.

ABSTRACT

Souza, V. P.; The role of Protein Structural Networks in the properties of a Beta-Glycosidase. 2017.303p. PhD Thesis - Graduate Program in Biochemistry. Instituto de Química, Universidade de São Paulo, São Paulo.

Analysis of protein structures as networks has been shown a powerful tool to understand their properties and to identify important residues. In the network analysis, residues that interact with each other are called connected. Some residues are essential to shorten the connection pathways between distant residues in the protein structure, being called central. Central residues have been proposed to have important roles in catalysis, thermal stability and allostery. In order to experimentally assess the correlation between the residue centrality and its importance in the protein properties, we use two approaches (chapter 1): The first one is to make single mutations at the central residues of a beta-glucosidase Sf β gly, changing those residues to alanine. The second one is to perturb a central residue (F251) by changing its environment through single mutations that introduces voids or additional volume. Next, we evaluate how those mutations affect the protein thermostability and function.

In general, we have observed (chapter 2) that mutations at central residues reduce the T_m in 2 - 15°C and increase the unfolding rate up to 20 times, suggesting that damages in the central residues make the protein more unstable. Moreover, we have observed (chapter 3) that the perturbation of the central residues reduces Sf β gly catalysis, which seems to arise from the same cause that lead to the loss of thermal stability. Besides that, in chapter 4, the investigation of oligomeric state of Sf β gli using SAXS indicated that this protein is mainly a dimer in 100 mM citrate-phosphate pH 6,0, whereas it forms large oligomers, possibly dodecamers, in 10 mM phosphate pH 6,0. In parallel it was shown that Sf β gly undergoes an activation process in 10 mM phosphate and its kinetic parameters converge

to those observed for Sfβgly in 100 mM citrate-phosphate. Protein Structural Networks were built considering also that there are links between the polypeptidic chains of the Sfβgly oligomers. We observed 5 residues that are central in all kind of oligomeric structures here analyzed. Three of these residues, E187, P188 and N329, play important roles in the catalysis of this enzyme, and two of them (S247 and N249 are described in this thesis. Lastly, in the chapter 5, we observed that central residues by closeness, betweenness and $C_{\Delta Lp}$ are concentrated at the top of the beta-barrel (C-terminal end of the beta-strands and subsequent loops), suggesting that this region, where the active site is placed, is close, in terms of contacts, to the whole Sfβgly structure. Moreover, we have built the Protein Structural Network of 21 beta-glucosidases of the Glucoside Hydrolases family 1, revealing that the closeness central residues are highly conserved, being located in the active site of these enzymes. On the other hand, betweenness central residues are located in the same sites in the structure of different beta-glucosidases, but they are not always conserved. Shortly, these data experimentally support the hypothesis that the residue centrality in Protein Structural Network is correlated with the protein properties, as catalysis and stability.

Keywords: Protein Structural Networks; enzymes, thermal stability; beta-glucosidases; small-angle x-ray scattering.

LISTA DE ABREVIATURAS

Abs – Absorbância

C – “Clustering coefficient”

C_{betweeness} – Parâmetro de centralidade por “*betweeness*”

C_{closeness} – Parâmetro de centralidade por “*closeness*”

CD – Técnica de Dicroísmo Circular

CF100 – Tampão citrato-fosfato 100 mM pH 6,0

C_p – “Clustering coefficient” característico da rede

C_{ΔLp} – Parâmetro de centralidade calculado pela mudança do caminho mínimo da rede com a deleção do resíduo.

DLS – Técnica de Espalhamento de Luz Dinâmico

DNA – Ácido desoxirribonucleico

dNTP – 2’deoxi-nucleotídeos 5’ trifosfatados

DSF – Técnica de “Differential Scanning Fluorimetry”

EC – “Enzyme Comission Number”

EDTA – Ácido etilenodiamino tetra-acético

EOM – “Ensemble Optimization Method”

F10 – Tampão fosfato de potássio 10mM pH 6,0.

FT – “Flow-Through” (amostra que não ligou na resina)

GH – Glisocídeo Hidrolases

HEWL – Lisozima de clara de ovo de galinha

I – Cultura de células após indução por adição de IPTG.

IPTG – Isopropil β-D-1-tiogalactopiranosideo

K_{SV} – Constante de Stern-Volmer

L – Menor caminho para ir de um nó a outro

LB – Meio de cultura “Luria Broth”

LFER – “Linear Free Energy Relationships”

L_p – Caminho mínimo da rede

M_w – Massa molecular

ND – Não determinado

NI – Cultura de células antes da indução por IPTG

Ni-NTA – Resina de afinidade de níquel carregado

PCR – “Polymerase Chain Reaction”

PDB – “Protein Data Bank”

PEG - polietilenoglicol

*p*NPFuc – *p*-nitrofenil- β -D-Fucopiranosídeo

*p*NPGal – *p*-nitrofenil- β -D-Galactopiranosídeo

*p*NPGLu – *p*-nitrofenil- β -D-Glucopiranosídeo

R_g – Raio de Giro

RMN – Ressonância Magnética Nuclear

Rpm – Rotações por minuto

SANS – Técnica de Espalhamento de Nêutrons a baixos ângulos

SAXS – Técnica de Espalhamento de Raios-X a baixos ângulos

SCA – “Statistical Coupled Analysis”

SDS-PAGE – Eletroforese desnaturante utilizando SDS

SDS – Dodecil sulfato de sódio

Sf β gli – Beta-glicosidase de *Spodoptera frugiperda*

TAE – Tampão TRIS-Acetato EDTA.

T_m – temperatura de *melting*

TRIS – Tris(hidroximetil)aminometano

TSA – “Thermal-shift assay”

SUMÁRIO

INTRODUÇÃO GERAL	25
A MATEMÁTICA DAS REDES DE ESTRUTURA PROTEICA	29
1.1 Introdução.....	29
1.1.1 Redes na análise de sistemas biológicos	29
1.1.2 Breve histórico sobre as redes e o desenvolvimento das Redes Estruturais Proteicas	30
1.1.3 Modelo “Mundo Pequeno” proposto por Watts & Strogatz (1998)	31
1.1.4. Modelo “Sem Escala” de Albert e Barabasi (2000)	35
1.1.5 Centralidade dos resíduos para as redes reais	39
1.1.6 Centralidade e função em proteínas	42
1.1.7 Beta-glicosidase como modelo experimental.....	44
1.1.8 Relacionando centralidade com propriedades da Sfβgli	46
1.2 Materiais e Métodos.....	46
1.2.1 Produção de uma Rede de Estrutura de Proteínas para Sfβgli	46
1.2.2 Linhagens de bactérias utilizadas no projeto	47
1.2.3 Construções pET46_Sfβgli selvagem e mutantes de resíduos centrais	48
1.2.4 Construções dos vetores contendo Sfβgli mutantes.....	49
1.2.5 Preparação de bactérias competentes.....	51
1.2.6 Transformação de bactérias <i>XL10 gold</i> por choque térmico	51
1.2.7 Transformação de bactérias BL21 gold(DE3), XL1 Blue e Nova Blue (DE3).....	52
1.2.8 Preparação de DNA plasmidial em maior escala	52
1.2.9 Análise dos vetores que codificam Sfβgli mutantes	53
1.2.10 Expressão de Sfβgli selvagem e mutantes.....	53
1.2.11 Lise de bactérias e purificação de Sfβgli selvagem e mutantes	54
1.2.12 SDS-PAGE	56
1.2.13 Determinação da concentração de Sfβgli.....	57

1.2.14	Coleta de espectros de fluorescência de triptofano	57
1.2.15	Apagamento da fluorescência dos triptofanos	58
1.2.16	Coleta de espectros de Dicroísmo Circular.....	58
1.3	Resultados e Discussão	59
1.3.1	Planejamento das mutações em Sfβgli para teste da correlação centralidade e função proteica ...	59
1.3.2	Construção das Sfβgli mutantes	63
1.3.3	Expressão e purificação de Sfβgli selvagem e mutantes	68
1.3.4	Espectro de fluorescência de triptofanos da Sfβgli selvagem e mutantes	71
1.4.5	Dicroísmo Circular (CD) das Sfβgli selvagem e mutantes	81
1.4	Conclusão.....	86
TERMOESTABILIDADE E AS REDES DE ESTRUTURA PROTEICA.....		89
2.1	Introdução.....	89
2.1.1	Termoestabilidade de Proteínas	89
2.1.2	Termoestabilidade e Redes de Estrutura Proteica	90
2.2	Materiais e Métodos.....	92
2.2.1	Determinação da Temperatura de <i>melting</i> (T_m) por Dicroísmo Circular (CD)	93
2.2.2	<i>Differential Scanning Fluorimetry</i> (DSF)	94
2.2.3	Cinética de desnaturação utilizando o corante <i>SyproOrange</i> ®	95
2.2.4	Plote de Eyring e cálculo da energia de ativação da desnaturação	96
2.2.5	Cinética de inativação acompanhada pela atividade enzimática remanescente	96
2.3	Resultados e Discussão	97
2.3.1	Determinação da T_m das Sfβgli selvagem e mutantes.....	97
2.4	Conclusões	138
CATÁLISE E AS REDES DE ESTRUTURA PROTEICA.....		141
3.1	Introdução.....	141
3.1.1	Glicosídeo Hidrolases – uma classe de enzimas com diversos papéis	141

3.1.2	Beta-glicosidases e sua importância	142
3.1.3	Bases moleculares do funcionamento das beta-glicosidases.....	144
3.1.4	Redes de Estrutura Proteica e catálise enzimática	148
3.2	Materiais e Métodos.....	150
3.2.1	Ensaio de atividade enzimática acompanhando formação de <i>p</i> -nitrofenol	150
3.2.2	Ensaio de atividade enzimática utilizando celobiose como substrato	151
3.2.3	Ensaio de cinética enzimática	152
3.2.4	Ensaio para determinação do efeito do pH na atividade enzimática.....	152
3.3	Resultados e Discussão	153
3.3.1	Efeitos na eficiência catalítica da Sfb β gli promovidos por mutações na vizinhança de um resíduo central da rede estrutural	153
3.3.2	Cálculo de mudança energética de catálise provocada pelas mutações na vizinhança de um resíduo central da rede estrutural.....	164
3.3.3	Mudanças de especificidade provocadas pelas mutações na vizinhança de um resíduo central da rede estrutural	169
3.3.4	Efeito das mutações na vizinhança de um resíduo central no pH ótimo da Sfb β gli	175
3.3.5	Correlação entre termoestabilidade e mudanças no sítio ativo	178
3.4	Conclusões	181
ESTUDO EXPERIMENTAL DO ESTADO OLIGOMÉRICO DA SFBGLI E PERSPECTIVAS UTILIZANDO A ANÁLISE POR REDES DE ESTRUTURA PROTEICA		183
4.1	Introdução.....	183
4.1.1	Importância do estado oligomérico de enzimas	183
4.1.2	Técnicas para caracterização de estados oligoméricos de proteínas	183
4.1.3.	Princípios da técnica de espalhamento de Raios-X a baixos ângulos (SAXS)	186
4.1.4	Estados oligoméricos de beta-glicosidases GH1.....	192
4.1.5	Uso de redes para avaliar estrutura quaternária de proteínas	194
4.2	Materiais e Métodos.....	195
4.2.1	Análise por Espalhamento de Luz dinâmico (DLS).....	195
4.2.2	Coleta de dados de espalhamento de raios-X a baixos ângulos.....	195

4.2.3	Transformada inversa de Fourier.....	196
4.2.4	Análise gráfica dos dados.....	196
4.2.5	Produção de um modelo utilizando o programa DAMMIF	197
4.2.6	Produção de curvas de espalhamento esperadas para estruturas em pdb.....	197
4.2.7	Modelagem de corpos rígidos utilizando o programa SASRef.	198
4.2.8	Análise dos dados utilizando o programa <i>Ensemble Optimization Method</i> (EOM)	199
4.2.9	Análise dos dados utilizando o programa <i>Oligomer</i>	200
4.2.10	Ensaio de cinética enzimática de SfbglI	200
4.2.11	Determinação da centralidade dos resíduos de diferentes estados oligoméricos da SfbglI	201
4.3	Resultados e Discussões.....	202
4.3.1	Determinação do estado oligomérico da SfbglI em solução	202
4.3.2	Análise por SAXS da SfbglI em tampão CF100	206
4.3.3	Análise por SAXS da SfbglI em tampão F10	219
4.3.4	Efeito do estado oligomérico na atividade de SfbglI	221
4.3.5	Análise de SAXS e de atividade enzimática da SfbglI em tampão F10 em diferentes temperaturas.....	226
4.3.6	Uso de redes para análise da estrutura quaternária da SfbglI e perspectivas.....	229
4.4	Conclusão.....	238
	REDES DE ESTRUTURA PROTEICA EM BETA-GLICOSIDASES.....	241
5.1	Introdução.....	241
5.1.1	Construindo Redes de Estrutura Proteica	241
5.1.2	Redes de Estrutura Proteica e enzimas	245
5.2	Materiais e Métodos.....	246
5.2.1	Construção da Rede de Estrutura de SfbglI e cálculo da centralidade C_{pLP}	246
5.2.2	Produção das redes utilizando o aplicativo <i>CMView</i> 1.1.1.....	246
5.2.3	Análise das redes de estrutura proteica utilizando o programa <i>Cytoscape</i>	248
5.3	Resultados e Discussão	248
5.3.1	Construção da rede de estrutura proteica utilizando diferentes critérios de conexão entre nós ..	248

5.3.2	Análise comparativa da rede estrutural de Sfβgli.....	256
5.3.3	Comparação de diversos tipos de centralidade na rede estrutural de Sfβgli	259
5.3.4	Análise de rede de beta-glicosidases da família GH1	269
5.4	Conclusões	288
CONCLUSÃO GERAL		291
ANEXOS.....		305

ÍNDICE DE FIGURAS

Figura 1. Representações de uma proteína (lisozima de clara de ovo; HEWL)..	30
Figura 2. Diferenças entre redes regulares, aleatórias e “Mundo Pequeno”	33
Figura 3. Diferenças entre a rede aleatória e a rede “Sem Escala”	36
Figura 4. Efeito de deleções em Redes do tipo Aleatória e Sem Escala..	38
Figura 5. Estrutura primária e terciária da beta-glicosidase de <i>Spodoptera frugiperda</i> (Sfβgli)..	45
Figura 6. Representação da construção pET46_Sfβgli	48
Figura 7. Rede de Estrutura Proteica da Sfβgli	61
Figura 8. Estratégias de perturbações dos resíduos centrais da Sfβgli.....	63
Figura 9. Eletroforese em gel de agarose do produto de reações de PCR.	64
Figura 10. Sequência de DNA completa que codifica a Sfβgli selvagem clonada no vetor pET46.	66
Figura 11. Sequenciamento de DNA do inserto que codifica Sfβgli selvagem e mutantes...	67
Figura 12. Análise da produção das Sfβgli selvagem e mutantes recombinantes em <i>BL21 gold (DE3)</i> através de SDS-PAGE.	69
Figura 13. Análise em SDSPAGE da purificação da Sfβgli selvagem e mutantes utilizando resina de Ni-NTA Agarose.....	70
Figura 14. Espectros de emissão de fluorescência de triptofanos da Sfβgli selvagem e mutantes.	76
Figura 15. Supressão da fluorescência dos resíduos triptofano de Sfβgli selvagem e mutantes utilizando acrilamida (Q).....	80
Figura 16. Valores de K_{sv} obtidos para Sfβgli selvagem e mutantes.	81
Figura 17. Espectro de CD para Sfβgli selvagem e mutantes.....	86

Figura 18. Perfil de desnaturação térmica da Sfβgli selvagem acompanhado por CD.....	98
Figura 19. Desnaturação térmica dos mutantes acompanhada através de CD. Leituras de elipsidade molar foram coletadas em 222 nm..	101
Figura 20. Comparação entre dos perfis obtidos por DSF das Sfβgli mutantes com a selvagem.	106
Figura 21. Curvas de DSF de todos os mutantes de Sfβgli.	110
Figura 22. T_m das Sfβgli selvagem e mutantes.	111
Figura 23. Vizinhança do resíduo F251.....	114
Figura 24. Desnaturação térmica da Sfβgli selvagem e do mutante A264F..	115
Figura 25. Efeito da mutação A264F sobre o resíduo central F251.....	116
Figura 26. Perfil cinético da desnaturação térmica a 45°C da Sfβgli selvagem e mutantes.....	121
Figura 27. Constantes cinéticas de desnaturação térmica (k_{desn}) a 45 °C das Sfβgli selvagem e mutantes..	122
Figura 28. Plotes de Eyring da desnaturação térmica da Sfβgli selvagem e mutantes. Experimentos foram realizados na faixa de 30 °C a 50 °C.	126
Figura 29. Perfil de energia dos mutantes durante o processo de desnaturação.....	128
Figura 30. Inativação térmica a 45 °C das Sfβgli selvagem e F251A.	129
Figura 31. Inativação térmica das Sfβgli selvagem e mutantes a 45°C.....	132
Figura 32. Mecanismo de catálise das beta-glicosidases GH1.....	144
Figura 33. Representação do sítio ativo de beta-glicosidases.	146
Figura 34. Curva-padrão de <i>p</i> -nitrofenolato	150
Figura 35. Curva padrão de glicose utilizando o reagente de detecção Glicose PP, da Analisa.	151
Figura 36. Ensaio de atividade enzimática das Sfβgli A264F e V336F.	156

Figura 37. Efeito da concentração do substrato <i>p</i> NP <u>Glu</u> na velocidade inicial da reação catalisada pela enzima S <u>f</u> β gli selvagem e mutantes.	158
Figura 38. Efeito da concentração de substrato <i>p</i> NP <u>Fuc</u> na velocidade inicial da reação catalisada pela S <u>f</u> β gli selvagem e mutantes.	159
Figura 39. Efeito da concentração de substrato <i>p</i> NP <u>Gal</u> na velocidade inicial da reação catalisada pela S <u>f</u> β gli selvagem e mutantes.	160
Figura 40. Efeito da concentração do substrato celobiose na velocidade inicial da reação catalisada pela S <u>f</u> β gli selvagem e mutantes.	161
Figura 41. Perfil energético de uma reação catalisada por enzima.	165
Figura 42. Efeito de uma mutação no perfil energético da reação catalisada por uma enzima.	166
Figura 43. Efeito das mutações no perfil energético da reação catalisada por S <u>f</u> β gli.	168
Figura 44. Estrutura dos substratos utilizados para a análise da especificidade das S <u>f</u> β gli selvagem e mutantes.	170
Figura 45. Valores de k_{cat}/K_m relativo da S <u>f</u> β gli selvagem e mutantes. wt, selvagem.	171
Figura 46. Análise LFER comparando as S <u>f</u> β gli selvagem e mutantes.	174
Figura 47. Comparação do efeito do pH na atividade enzimática das S <u>f</u> β gli selvagem e mutantes.	176
Figura 48. Relação entre a termoestabilidade e a eficiência catalítica da S <u>f</u> β gli selvagem e mutantes que perturbam o resíduo central F251. <i>gluco</i> corresponde à <i>p</i> NP <u>Glu</u>	179
Figura 49. Esquema representando a propagação do efeito das mutações na vizinhança de um resíduo central (<i>hub</i>) pela proteína chegando ao sítio-ativo.	181
Figura 50. Semelhanças e diferenças entre as técnicas de cristalografia de raios-X e SAXS.	185
Figura 51. Esquema representando o espalhamento de raios x de uma partícula.	186
Figura 52. Vetor da onda espalhada por uma partícula.	187
Figura 53. Esquema geral de um experimento de SAXS.	188

Figura 54. Perfil de espalhamento esperado para partículas com diferentes formas geométricas.	189
Figura 55. Exemplo de plotes de Guinier que desviam da linearidade.....	191
Figura 56. Estrutura cristalográfica da Sfβgli (5CG0).....	202
Figura 57. Perfis de espalhamento dinâmico de luz (DLS) da Sfβgli selvagem em tampão CF100 e F10.	204
Figura 58. Perfil de espalhamento por SAXS da Sfβgli selvagem, a 25°C e pH 6,0.	205
Figura 59. Plote de Guinier obtido para a Sfβgli selvagem em tampão CF100 (preto) e F20 (vermelho).....	205
Figura 60. Curva de espalhamento de raios-X a baixos ângulos. (.....	208
Figura 61. Relação entre a intensidade quando o ângulo tende a 0 - $I(0)$ – no plote de Guinier e a concentração de Sfβgli. A linha em preto é a regressão linear dos dados.	209
Figura 62. Comparação entre o espalhamento experimentalmente observado para a Sfβgli em CF100 e os perfis esperados para diferentes formas oligoméricas da enzima.	209
Figura 63. Curva da função dos pares de distância $[p(r)]$ das amostras de Sfβgli em diferentes concentrações, todas em tampão CF100. Dados obtidos através do programa GNOM.	210
Figura 64. Função dos pares $[p(r)]$ de diversos estados oligoméricos da Sfβgli.	211
Figura 65. Duas perspectivas do modelo <i>ab initio</i> gerado pelo programa DAMMIF em sobreposição com a estrutura cristalográfica de Sfβgli.	212
Figura 66. Modelo de um dímero da Sfβgli.	213
Figura 67. Perfil de espalhamento de modelos de dímero de Sfβgli obtidos a partir da estrutura cristalográfica (curva preta) e pelo programa DIMFOM (curva vermelha).	214
Figura 68. Comparação entre o plote de Krakty do modelo dimérico da Sfβgli com o plote produzido a partir dos dados experimentais.	215
Figura 69. Modelos de dímeros de Sfβgli gerados pelo programa EOM.	217

Figura 70. Fração volumétrica de dímeros e tetrâmeros de Sfβgli em tampão CF100 estimada pelo programa OLIGOMER.	218
Figura 71. Comparação entre o perfil de espalhamento calculado pelo CRY SOL para diferentes formas oligoméricas de Sfβgli e o espalhamento experimental obtido com a enzima 0,8 mg/mL em F10.	220
Figura 72. Parâmetros cinéticos da Sfβgli selvagem em tampão CF100 em diferentes dias após a purificação incubação a 4°C.	223
Figura 73. Parâmetros cinéticos da Sfβgli selvagem em tampão F10 em diferentes dias após a purificação incubação a 4°C.	224
Figura 74. Esquema da ativação de Sfβgli a 4 °C em tampão F10.	226
Figura 75. Espalhamento por SAXS da Sfβgli em tampão F10 obtido em diferentes temperaturas.	228
Figura 76. Plote de Eyring da atividade da Sfβgli obtida em diferentes temperaturas reacionais.	229
Figura 77. Dependência do caminho mínimo médio da rede (L) com relação ao seu tamanho (N; número de nós) para diferentes modelos (Regular, Randômico e Mundo Pequeno). .	231
Figura 78. Resíduos centrais por <i>closeness</i> no monômero (em vermelho), dímero (em azul) e tetrâmero (amarelo) da Sfβgli.	233
Figura 79. Diferença de centralidade por <i>closeness</i> observada para o resíduo n analisando a transição de monômero para o dímero da Sfβgli (gráfico superior) e de dímero para tetrâmero (gráfico inferior).	234
Figura 80. Resíduos centrais por <i>betweeness</i> em diferentes estados oligoméricos da Sfβgli.	236
Figura 81. Diferenças de centralidade por <i>betweeness</i> observadas para o resíduo n ao analisar as transições de monômero para dímero (gráfico superior) e monômero para tetrâmero (gráfico inferior)..	237
Figura 82. Produção da rede de estrutura proteica da Sfβgli considerando a distância entre resíduos como critério para formação de contatos.....	247

Figura 83. Comparação entre redes de estrutura proteica produzidas a partir do modelo cristalográfico da Sfβgli (5CG0) utilizando diferentes definições de contatos entre resíduos.	249
Figura 84. Redes de estrutura proteica da Sfβgli produzidas a partir da estrutura cristalográfica 5CG0 utilizando diferentes distâncias de corte para definição de contatos entre resíduos.	252
Figura 85. Parâmetros das redes estruturais da Sfβgli em função da distância de corte para contato entre resíduos.	255
Figura 86. Características “Mundo Pequeno” das redes de estrutura proteica de Sfβgli... ..	257
Figura 87. Representação dos resíduos centrais da Rede de Estrutura Proteica da Sfβgli.	261
Figura 88. Distribuição da centralidade dos nós da rede estrutural de Sfβgli utilizando diferentes critérios.....	264
Figura 89. Representação da estrutura secundária e da rede estrutural da Sfβgli produzidos a partir da sua estrutura cristalográfica (5CG0).....	265
Figura 90. Representação dos resíduos centrais na estrutura da Sfβgli.....	267
Figura 91. Relação entre o parâmetro de centralidade por <i>closeness</i> e o <i>B-factor</i> de cada um dos resíduos de Sfβgli.....	269
Figura 92. Representação circular das redes de estrutura proteica de beta-glicosidases GH1.	275
Figura 93. Alinhamento de sequencias de beta-glicosidases e comparação dos resíduos centrais por grau.....	278
Figura 94. Alinhamento de sequências de diferentes beta-glicosidases GH1 com destaque para os resíduos centrais por <i>betweeness</i>	283
Figura 95. Alinhamento de sequências de diferentes beta-glicosidases GH1 com destaque para os resíduos centrais por <i>closeness</i>	287

ÍNDICE DE TABELAS

Tabela 1. Iniciadores mutagênicos utilizados nos experimentos de mutagênese sítio-dirigida para produção das Sfβgli mutantes.	50
Tabela 2. Concentração das Sfβgli selvagem e mutantes recombinantes após a purificação.....	72
Tabela 3. Valores de T_m da Sfβgli selvagem e mutantes determinados através de dicroísmo circular (CD) e ThermoFluor. Extraído de Souza <i>et al.</i> (2016).	112
Tabela 4. Parâmetros termodinâmicos da desnaturação térmica da Sfβgli selvagem e mutantes a 45°C..	127
Tabela 5. Constantes de velocidade de desnaturação e inativação térmica das Sfβgli selvagem e mutantes.	137
Tabela 6. Valores de K_m (em mM) para 4 tipos de substrato obtidos para a Sfβgli selvagem e mutantes	162
Tabela 7. k_{cat} (s^{-1}) das Sfβgli selvagem e mutantes para hidrólise de 4 tipos de substrato.....	163
Tabela 8. k_{cat}/K_m ($mM^{-1}s^{-1}$) das Sfβgli selvagem e mutantes para 4 tipos de substrato.	163
Tabela 9. $\Delta\Delta G^\ddagger$ (kJ/mol) entre diferentes substratos para as Sfβgli selvagem e mutantes.	172
Tabela 10. Valores de pK_a dos resíduos catalíticos e pH ótimo das Sfβgli selvagem e mutantes.	177
Tabela 11. Resíduos centrais por <i>closeness</i> da Rede de Estrutura Proteica produzida a partir da estrutura cristalográfica da Sfβgli (5CG0) utilizando diferentes critérios de contato.	250
Tabela 12. Relação de resíduos centrais por <i>closeness</i> em redes estruturais de Sfβgli produzidas com diferentes distâncias de corte como critério de estabelecimento de conexões entre resíduos.....	255
Tabela 13. Relação de resíduos centrais na Rede de Estrutura da Sfβgli segundo diferentes parâmetros.	262
Tabela 14. Parâmetros da rede de estrutura proteica de diferentes beta-glicosidases GH1.....	270

INTRODUÇÃO GERAL

Proteínas são biopolímeros contendo centenas de aminoácidos. No entanto, nem todos os resíduos de uma proteína são igualmente importantes em suas propriedades, sendo que mutações em algumas posições parecem afetar fortemente determinadas propriedades, enquanto em outras posições podem não causar nenhum efeito mensurável (Capra e Singh, 2007). Mesmo assim, o papel que cada um desses aminoácidos possui na estrutura e função da proteína não é simples de determinar. A correlação entre os resíduos e a sua importância para as propriedades das proteínas, além de ser uma questão em aberto da bioquímica, também é de extremo interesse para a engenharia de proteínas.

Muitas abordagens têm sido utilizadas para determinar os resíduos que desempenham funções importantes em proteínas. Uma das abordagens mais comuns é o alinhamento da sequência proteica entre diferentes proteínas homólogas. Como aminoácidos importantes para estrutura e função de proteínas tendem a ser conservados devido à pressão evolutiva, é possível identificar estes resíduos pelo seu grau de conservação e comparação com resíduos importantes de outras proteínas homólogas já caracterizadas. No entanto, essa análise exige que a proteína de interesse possua diversas proteínas análogas com sequências conhecidas, e também a caracterização de algumas delas. Além disso, é necessário que a pressão evolutiva crie um alto grau de conservação deste resíduo, o que nem sempre ocorre. Um exemplo deste fenômeno é relatado por Suemori (2013), que através de mutações identificou que resíduos conservados e não conservados desempenham papéis importantes na estrutura e função da enzima p-hidroxibenzoato hidroxilase.

Em uma abordagem mais recente, Ranganathan e colaboradores analisaram proteínas avaliando a covariância da frequência de ocorrência de diferentes aminoácidos ao longo da estrutura primária de uma proteína. Para isso, os autores desenvolveram a técnica chamada de “Statistical Coupled Analysis (SCA)”, onde não é avaliado o grau de conservação de resíduos, mas a influência que a presença de um determinado resíduo provoca na conservação de outro resíduo da estrutura. Assim, esta abordagem permite identificar conjuntos de resíduos que se influenciam mutuamente. Através dessa análise, os autores determinaram grupos de resíduos que formam rotas de propagação de efeitos alostéricos dentro das estruturas protéicas (Lockless e Ranganathan, 1999; Süel *et al.*, 2003). De qualquer modo, para realizar essa análise, ainda é necessária a existência de centenas de proteínas homólogas com sequências conhecidas.

Em 2001, Vendruscolo e colaboradores inauguraram uma nova abordagem para análise de estrutura e função em proteínas, analisando estruturas de pequenas proteínas como redes de elementos interconectados. Nestas redes os resíduos de aminoácidos são os nós e as interações não-covalentes e covalentes entre eles são as conexões da rede. Isto permitiu empregar análises matemáticas que identificam os nós essenciais para a conectividade da rede, e assim uma propriedade chamada centralidade foi atribuída a cada resíduo da estrutura proteica. Os autores sugerem, neste trabalho, que os resíduos centrais da rede da estrutura proteica no estado de transição do processo de enovelamento são resíduos importantes para o seu dobramento, agindo analogamente a pontos de nucleação (Vendruscolo *et al.*, 2001; Vendruscolo, Manu *et al.*, 2002). A partir disso, diversos trabalhos têm modelado a estrutura de proteínas como redes. Nestes trabalhos tem sido

observado que os resíduos centrais das Redes de Estrutura Proteica (do inglês *Protein Structural Networks*) possuem papéis importantes na função dessas proteínas (Amitai *et al.*, 2004; Del Sol *et al.*, 2006; Böde *et al.*, 2007; Li *et al.*, 2011; Niknam *et al.*, 2016; Mallik e Kundu, 2017), na dinâmica e alosteria (Atilgan *et al.*, 2004; Del_Sol *et al.*, 2006; Szalay e Csermely, 2013; Tse e Verkhivker, 2015; Fokas *et al.*, 2016) e na estabilidade de proteínas (Brinda e Vishveshwara, 2005; Vijayabaskar e Vishveshwara, 2010).

Contudo, esses trabalhos foram feitos utilizando proteínas previamente caracterizadas, e dessa forma não foram desenhados para avaliar experimentalmente a correlação entre a centralidade dos resíduos na rede estrutural e a sua importância em propriedades das proteínas.

Portanto, esta tese tem como objetivo geral avaliar experimentalmente a correlação entre a centralidade de resíduos na rede de estrutura proteica com a sua importância nas propriedades de uma proteína, a beta-glicosidase de *Spodoptera frugiperda* (Sfβgli). Desta forma, no capítulo 1 desta tese será discutido o planejamento experimental para avaliar essa correlação, englobando a produção e purificação das Sfβgli mutantes e selvagem que serão usadas como modelos experimentais. No capítulo 2, será feita uma análise entre a correlação da centralidade dos resíduos na rede estrutural com a termoestabilidade da Sfβgli. No capítulo 3 será analisada a correlação da centralidade dos resíduos com a sua importância na catálise da hidrólise pela beta-glicosídeos pela Sfβgli. No capítulo 4, será feita uma análise da estrutura quaternária da Sfβgli, investigando a correlação da centralidade dos resíduos com a sua importância na interação proteína-proteína. Por fim, no capítulo 5 desta tese, serão comparados diferentes parâmetros de centralidade dos

resíduos da rede de estrutura proteica da SfbgI e sua correlação com o grau de conservação dos resíduos centrais.

1

A MATEMÁTICA DAS REDES DE ESTRUTURA PROTEICA

1.1 Introdução

1.1.1 Redes na análise de sistemas biológicos

Redes vêm sendo cada vez mais utilizadas para analisar sistemas biológicos. Isso porque elas descrevem interações entre grandes conjuntos de entes e permitem analisar todas estas interações de forma global, o que é de fundamental interesse em sistemas complexos. Exemplos de aplicação de redes em sistemas biológicos são as redes metabólicas, redes de interação proteína-proteína, rede de transcrição, rede de sinalização e a rede gerada por árvores filogenéticas (Khanin e Wit, 2006; Bachmaier *et al.*, 2014).

Dessa forma, para compreender melhor a estrutura de proteínas, também é possível modelá-la como uma rede, considerando que os aminoácidos que a compõe são os nós da rede e as interações covalentes e não covalentes entre estes resíduos são as conexões, conforme representado na Figura 1.

Figura 1. Representações de uma proteína (lisozima de clara de ovo; HEWL). Em A, representação da estrutura terciária pelo modelo de cartoon. Em B, representação da estrutura da proteína como uma rede, sendo que as esferas representam os aminoácidos e os traços, as interações entre eles. Interações entre aminoácidos são definidas com base em uma distância máxima de separação.

Assim, é possível analisar a estrutura proteica considerando que todas as interações de um dado resíduo são igualmente importantes como representado na Figura 1B, sendo eles vizinhos em sequência ou não (Greene, 2012). Além de permitir uma análise global da estrutura proteica, a linguagem de redes permite compreender melhor a organização das interações contidas nas proteínas.

1.1.2 Breve histórico sobre as redes e o desenvolvimento das Redes Estruturais Proteicas

Redes têm sido utilizadas para resolver problemas por pelo menos 400 anos. Um dos primeiros registros do uso de redes para resolver um problema real foi o de Euler, em 1736. Ainda que sem a representação da rede como conhecemos hoje (Kruja *et al.*, 2002), Euler utilizou princípios fundamentais das redes para resolver o “Problema das Pontes de Königsberg”, prevendo que não seria possível sair de um ponto da cidade de Königsberg,

atravessar todas as suas pontes e retornar ao local de origem sem cruzar por duas vezes a mesma ponte.

Desta forma, apesar de redes começarem a ser desenvolvidas no século XVIII, o uso de redes para resolver problemas reais ficou restrito a pequenos sistemas (que podiam ser desenhados) até a década de 50. Foi nos anos 50 que Erdős e Renyi produziram o primeiro modelo para descrever as redes grandes, o chamado “Modelo Aleatório”, produzido criando conexões de forma aleatória entre os nós da rede. Apesar de este modelo ter trazido notáveis avanços na teoria das redes (ou grafos) como a definição de diversos parâmetros até hoje utilizados para caracterizar as redes, o que foi observado é que o modelo proposto por Erdős e Renyi tinha falhas ao explicar a maior parte das redes reais. Com isso, novos modelos foram desenvolvidos para descrever melhor as grandes redes reais (Barabasi, 2009).

1.1.3 Modelo “Mundo Pequeno” proposto por Watts & Strogatz (1998)

Um modelo que atualmente é bastante utilizado para descrever as redes reais é o modelo “Mundo-Pequeno” (Watts e Strogatz, 1998). Ele foi introduzido considerando que as redes reais são híbridas, ou seja, não são completamente regulares e nem totalmente aleatórias (Figura 2A). Isso significa que parte das conexões das redes reais normalmente é regular, *i.e.* todos os nós têm exatamente o mesmo número de conexões, gerando um padrão repetitivo na rede, enquanto que outra parte é organizada de forma aleatória.

Para compreender a regularidade da rede, os autores introduziram um parâmetro chamado de “*Clustering Coefficient*” (C), que pode ser calculado para cada um dos nós.

Para compreender o significado físico desse parâmetro, podemos considerar um nó i da rede, que faz contato com outros nós, dentre eles j e k . O parâmetro C_i é a probabilidade de que j e k estejam conectados, ou seja, de que os vizinhos do nó i estejam conectados entre si. Para caracterizar a rede como um todo com relação este parâmetro, podemos calcular o parâmetro C para cada um dos nós da rede, e fazer uma média destes valores, encontrando o parâmetro característico da rede C_p . Ao caracterizar redes utilizando este parâmetro, Watts e Strogatz (1998) observaram que as redes reais possuem valores de C_p maiores do que o previsto pelo modelo aleatório de Erdős e Renyi, indicando a existência de grupos de nós altamente conectados entre si e pouco conectados com nós fora do grupo (Figura 2B).

Outro parâmetro analisado pelos autores foi o parâmetro de caminho mínimo (L). O caminho mínimo entre dois nós i e j é calculado determinando o menor número de conexões necessárias para ligar i e j . Desta forma, considerando combinatoriamente todos os pares possíveis de nós, é possível calcular o caminho mínimo entre todos os pares possíveis, e fazer uma média entre os valores de caminho mínimo, a qual chamamos de L_p . Ao avaliar as redes reais, os autores observaram que os valores de L_p são bastante baixos, indicando que um pequeno número de conexões é necessário para conectar quaisquer nós da rede (Figura 2B). Essa propriedade é popularmente conhecida como “seis graus de separação”, uma ideia até então não científica de que quaisquer duas pessoas do mundo estão somente 6 passos distantes uma das outras (Barabasi, 2009).

Figura 2. Diferenças entre redes regulares, aleatórias e “Mundo Pequeno”. Em A está ilustrado o modelo “Mundo Pequeno”, gerado considerando uma rede completamente regular (como à esquerda) e refazendo algumas conexões de forma aleatória (como à direita). O número de conexões que são modificadas e tornadas aleatórias é determinado pela probabilidade p de um nó conectar qualquer outro nó da rede, representada na seta abaixo das redes. Em B, um gráfico que relaciona a probabilidade p com os parâmetros C_p e L_p . Como indicado na figura, redes altamente regulares ($p=0$) possuem um alto valor de C_p e alto valor de L_p . Redes aleatórias ($p=1$) possuem baixos valores de C_p e L_p . Já as redes “Mundo Pequeno”, englobadas entre essas faixas de p , possuem alto valor de C_p e baixo valor de L_p . Figura adaptada de Watts e Strogatz (1998).

Com isso, os autores observaram que o modelo “Mundo Pequeno” descreve bem a maior parte das redes reais, pois o modelo gera redes com altos valores de C_p e baixos valores de L_p , conforme indicado na Figura 2B. Além disso, neste tipo de rede, ao plotar um

histograma do número de conexões de cada nó (grau), observa-se uma distribuição de Poisson, que ocorre em parte das redes reais.

Já ao avaliar Redes de Estrutura Proteica, Kannan e Vishveshwara (1999) observaram a presença de *clusters* aromáticos, sendo que proteínas mais termoestáveis possuem *clusters* aromáticos adicionais. Com isso, a “*clusterização*” (agregação), que é um fenômeno observado em redes com propriedades “Mundo Pequeno”, também é encontrada em Redes de Estrutura Proteica. Além disso, Greene e Higman (2003), e Bagler e Sinha (2005) observaram que Redes de Estrutura Proteica geradas a partir de proteínas de diferentes topologias têm similaridades ao modelo “Mundo Pequeno”. Isso porque parte das conexões ocorre entre resíduos próximos na estrutura primária e parte entre resíduos distantes em pelo menos 10 posições, o que gera uma combinação de padrões regulares e aleatórios. Além disso, o histograma do grau dos nós é bem descrito por uma distribuição de Poisson, o que é compatível com o modelo “Mundo Pequeno”. Aftabuddin e Kundu (2007) também observaram a adequação do modelo “Mundo Pequeno” para descrever redes contidas dentro das Redes de Estrutura Proteica, indicando que as subredes de interações hidrofóbicas, hidrofílicas e de resíduos carregados também se comportam de acordo com o modelo “Mundo Pequeno”.

No entanto, alguns grupos têm discordado de que o modelo “Mundo Pequeno” é o melhor modelo para descrever Redes de Estrutura Proteica. Morita e Takano (2009) indicaram que o melhor modelo para descrever as redes de estrutura proteica seria o de uma “Rede de Percolação”, uma subclasse contida dentro das chamadas “Redes Fractais”.

De fato, Leitner (2008) tem utilizado a similaridade das estruturas de proteínas a fractais para descrever a propagação da energia térmica pela estrutura proteica.

De qualquer modo, apesar destas divergências, diversos trabalhos têm apontado que as Redes de Estrutura Proteica têm “Propriedades de Mundo Pequeno”.

1.1.4. Modelo “Sem Escala” de Albert e Barabasi (2000)

Outro modelo de rede desenvolvido quase simultaneamente ao modelo “Mundo Pequeno” foi o “Sem Escala”. Neste modelo, proposto por Albert e Barabasi (2000), os nós mais conectados têm maior probabilidade de receberem novas conexões, seguindo uma propriedade que os autores chamam de “Ricos ficam cada vez mais ricos”. Analisando redes da *World-Wide-Web*, na qual os *websites* são os nós e os *links* entre eles são as conexões, os autores observaram que *sites* muito conectados são mais divulgados, e por isso têm maior probabilidade de receber novos *links*. Isso faz com que *sites* muito conectados sejam cada vez mais acessados, enquanto *sites* pouco conectados mantenham-se indefinidamente pouco acessados. Na linguagem das redes isso significa que nós muito conectados tendem a ser ainda mais conectados. Além disso, Barabasi também aponta a importância da dinâmica da formação dessas redes, pois na rede da *World-Wide-Web*, alguns *sites* são antigos e tendem a ter mais conexões, enquanto outros são muito recentes e acabam não tendo conexões por falta de tempo para que isso ocorra (Barabasi, 2009). Desta forma, o autor propõe que diversas redes são bem descritas pelo modelo “Sem Escala”, por também compartilhar destas características.

Uma forma de diferenciar redes do tipo “Sem Escala” de Redes Aleatórias é plotar um histograma com o número de conexões dos seus nós. Enquanto redes aleatórias possuem distribuição de Poisson, redes “Sem Escala” possuem distribuição que segue a lei das potências ($P = k^{-n}$, onde n é uma constante e k o número de conexões de um dado nó), conforme representado na Figura 3. Isso ocorre porque nesse tipo de rede, nós com número de conexões bem acima da média (muitas vezes, mais de 3 ou 4 desvios padrão da média) ocorrem em porcentagem maior do que o previsto pelo modelo aleatório.

Figura 3. Diferenças entre a rede aleatória e a rede “Sem Escala”. Extraído de Barabasi (2009).

Albert e Barabasi (2000) também observaram que a robustez da rede é mantida pelos nós mais conectados. Em uma análise computacional, os autores produziram uma

rede “Sem Escala” e fizeram um teste de robustez. Primeiramente, os autores apagaram aleatoriamente nós da rede e avaliaram como essas deleções afetavam a sua fragmentação e o seu parâmetro L_p . Foi notado que deleções aleatórias dos nós (falhas) pouco afetam o parâmetro L_p , e raramente provocam fragmentação da rede. Já em outra linha, foram removidos os nós seguindo a ordem decrescente do número de conexões (ataque). Os autores observaram que as deleções dos nós mais conectados (chamados de *hubs*) tem alta probabilidade de fragmentar a rede em conjuntos menores, além de aumentar drasticamente o parâmetro L_p . Isso indica que para, uma rede “Sem Escala” falhas aleatórias praticamente não afetam a rede como um todo, enquanto que ataques aos nós mais conectados provocam o colapso da rede. Esse fenômeno ocorre de forma diferente com Redes Aleatórias (Figura 4), onde falhas e ataques causam fragmentação e aumento de L_p do mesmo modo. Com isso, foi proposto que redes reais poderiam sofrer uma “pressão evolutiva” para se organizarem como “Sem Escala”, o que as garantiria robustez contra falhas aleatórias.

Nesta linha, Greene e Higman (2003) demonstraram computacionalmente que Redes de Estrutura Proteica possuem robustez, ou seja, resistem a falhas aleatórias, mas colapsam ao sofrer ataques em nós bem conectados. As autoras ainda discutem que apesar das Redes de Estrutura Proteica serem bem descritas pelo modelo “Mundo Pequeno”, existe dentro delas subredes com propriedades “Sem Escala”. Estas subredes podem ser produzidas considerando somente os contatos de longo alcance, ou seja, entre resíduos distantes em pelo menos 10 posições na estrutura primária. Ao considerar este tipo de subrede, a distribuição de graus de conexões é bem descrita por um gráfico

semilogarítmico, indicando que esta subrede carrega semelhanças funcionais ao modelo “Sem Escala”.

Figura 4. Efeito de deleções em Redes do tipo Aleatória e Sem Escala. Em A, valor de L_p em função do número de deleção de nós de Redes Aleatórias (A) e (LE). Em B, um esquema ilustrando a fragmentação de redes aleatória e Sem Escala em função de falhas ou ataques, segundo definido a seguir. As deleções podem ser ataques quando são direcionadas aos nós mais conectados da rede, ou falhas quando são feitas aleatoriamente. Adaptada de Albert e Barabasi (2000).

1.1.5 Centralidade dos resíduos para as redes reais

Em redes reais, especialmente as que possuem propriedades de “Sem Escala”, não há um padrão médio de nó e então alguns são considerados mais importantes para a rede do que outros. Estes nós são chamados de nós centrais da rede. Diversos algoritmos têm sido utilizados para calcular a centralidade dos nós das redes. Alguns desses parâmetros de centralidade são:

Centralidade em grau: Mede a centralidade de um nó considerando o número de conexões que este possui. Nesse sentido, nós muito conectados são chamados *hubs*. Assim, formalmente, o parâmetro de centralidade por grau do resíduo i , k_i , é o número de conexões que ele possui. Fokas e colaboradores (2016) apontaram que para Redes de Estrutura Proteica, resíduos *hubs* normalmente são encontrados em regiões mais internas das proteínas, e que devido ao alto número de contatos, tendem a ter uma taxa de evolução baixa.

Centralidade closeness: Mede-se a centralidade de um nó considerando o quão próximo este nó está dos demais em número de conexões. Para isso, o caminho mínimo L deste nó para todos os demais nós da rede é calculado, sendo que a centralidade é o inverso da média de todos os L medidos. Formalmente, o cálculo é feito através da Equação 1:

$$C_{Closeness}(i) = \frac{n}{\sum_{x \neq i} dist(i, x)}$$

Equação 1

Onde $C_{closeness}(i)$, o parâmetro de *closeness* do nó i ; corresponde a n , o numero total de pares de nós envolvendo o resíduo i dividido pela somatória das menores distâncias entre i e os demais nós da rede, x . Distâncias são medidas em número de conexões.

Centralidade betweeness: Determina-se a centralidade de um nó considerando o número de vezes que este nó em particular está inserido em um caminho mínimo entre dois pares quaisquer de nós da rede. Desta forma, considerando os caminhos mínimos L que ligam quaisquer dois nós da rede, um nó com alto valor de *betweeness* aparece frequentemente como parte desses caminhos mínimos. A definição do parâmetro de *betweeness* é dada pela Equação 2:

$$C_{Betweeness}(i) = \sum_{a \neq i \neq b} \frac{\sigma_{ab}(i)}{\sigma_{ab}} \quad \text{Equação 2}$$

Onde $C_{Betweeness}(i)$ é o parâmetro de *betweeness* do nó i , σ_{ab} é o número de caminhos mínimos do nó a até o nó b , e $\sigma_{ab}(i)$ é o número de caminhos mínimos de a até b que passam pelo nó i .

Centralidade por variação de L_p ($C_{\Delta L_p}$): del Sol e O'meara (2005) determinam a centralidade de resíduos de Redes de Estrutura Proteica removendo-os computacionalmente da estrutura e recalculando o valor de L_p após esta deleção. Os resíduos centrais são aqueles cuja deleção computacional provoca um aumento significativo do valor de L_p . Para calcular esta centralidade, inicialmente deve ser calculado o caminho mínimo médio da rede, L_p , através da Equação 3:

$$L_p = \frac{1}{n} \sum_{a \neq x} dist(a, x) \quad \text{Equação 3}$$

Onde L_p é o caminho mínimo médio da rede, n é o número total de nós, e $dist(a, x)$ é o caminho mínimo do nó a até o nó x medido em número de conexões.

Após este cálculo, o nó i e suas conexões são removidos computacionalmente da rede e a mudança no valor de L_p é calculada pela Equação 4:

$$\Delta L_i = |L_p - L_{rem.i}| \quad \text{Equação 4}$$

Sendo ΔL_i a variação de L_p devido a deleção do resíduo i , e $L_{rem.i}$ o valor de L_p após a remoção do nó i .

Finalmente é calculado o $\Delta L_{médio}$ e seu desvio padrão (σ), sendo então a centralidade do nó i expressa em termos de desvios padrões em relação ao $\Delta L_{médio}$ segundo a Equação 5:

$$C_{\Delta L_p} = \frac{\Delta L_i - \Delta L_{médio}}{\sigma} \quad \text{Equação 5}$$

Centralidade de perturbação: Proposta por Szalay e Csermely (2013), este tipo de centralidade é calculada promovendo uma perturbação na rede através de um programa desenvolvido pelo grupo, o programa *Turbine*. Para isso, a centralidade do nó i é calculada aplicando uma dada quantidade de energia neste nó e avaliando a velocidade de dissipação dessa energia através da rede (em quantos passos esta energia é dissipada para os demais

nós). No trabalho, os autores observam que há certa correlação entre os nós centrais identificados pelo critério de perturbação com aqueles identificados por *closeness*.

1.1.6 Centralidade e função em proteínas

Muitos trabalhos têm apontado que a centralidade dos resíduos pode estar associada a funções que estes resíduos desempenham na proteína. Del Sol e colaboradores (2006) analisando 7 famílias de proteínas empregando o conceito de centralidade por $C_{\Delta Lp}$ (mioglobinas, receptores acoplados a proteínas G, tripsinas, hemoglobinas, oligossacarídeo fosforilases, domínios de interação com ligantes em receptores nucleares e proteases retrovirais), perceberam que muitos resíduos centrais destas proteínas desempenham papéis fundamentais em rotas alostéricas, sugerindo que estes resíduos possuem papel de propagar a informação ao longo da estrutura proteica. Adicionalmente, Amitai e colaboradores (2004) indicam em seu trabalho que os resíduos constituintes do sítio ativo de proteínas frequentemente são centrais por $C_{closeness}$, e desta forma a análise de centralidade dos resíduos poderia ser uma ferramenta para identificar sítios ativos de enzimas. De fato, combinando este critério de centralidade com uma baixa acessibilidade ao solvente, Amitai e colaboradores (2004) encontraram sítios ativos de 70% de 178 estruturas proteicas previamente caracterizadas. Além disso, Brinda e Vishveshwara (2005) observaram que proteínas termofílicas possuem um maior número de resíduos centrais em grau do que suas análogas mesofílicas, indicando que a termoestabilidade das proteínas pode estar relacionada à organização da sua rede de estrutura proteica. Por fim, além destes trabalhos acima, existem dois trabalhos pioneiros que iniciaram a investigação das

proteínas como redes. Os autores desses trabalhos propõem que os resíduos centrais desempenham papéis fundamentais para o enovelamento de proteínas (Vendruscolo *et al.*, 2001; Vendruscolo, Manu *et al.*, 2002). Nestes trabalhos, Vendruscolo e colaboradores sugerem que a construção de uma rede utilizando a estrutura do estado de transição do enovelamento pode ser uma ferramenta para compreender o enovelamento. Os autores mostram utilizando pequenas proteínas que os resíduos centrais do estado de transição do enovelamento também são centrais na estrutura da forma nativa das proteínas, mas que nem todos os centrais da forma nativa são centrais do estado de transição. Dessa forma, eles sugerem que a centralidade dos resíduos também pode ter relação com o enovelamento das proteínas.

Enfim há diversos trabalhos que vêm utilizando as Redes de Estrutura Proteica para identificar resíduos e regiões importantes para função e propriedades de proteínas. No entanto, além deste projeto, não existem trabalhos desenhados para comprovar experimentalmente se os resíduos centrais da rede desempenham, de fato, papéis importantes na determinação de propriedades das proteínas. Adicionalmente observa-se na literatura uma lacuna de trabalhos que promovam perturbações em resíduos centrais e em resíduos não centrais, de forma a avaliar comparativamente se as perturbações nos resíduos centrais causam mudanças mais significativas nas propriedades das proteínas. Também há uma falta de trabalhos experimentais mostrando que perturbações dirigidas para a vizinhança de um resíduo central são propagadas por toda estrutura proteica, corroborando experimentalmente a hipótese da participação de resíduos centrais em vias

de contato que transmitem mais eficientemente alterações através da estrutura de proteínas.

1.1.7 Beta-glicosidase como modelo experimental

Para preenchimento das lacunas acima descritas é preciso buscar um modelo experimental que combine informações sobre estrutura primária e terciária, um grande número de proteínas homólogas e uma caracterização funcional detalhada. Deste modo, seria possível elaborar a representação em rede e a identificação dos resíduos centrais e combinar estas informações com o mapeamento estrutural e funcional destes resíduos. Com base nestes argumentos foi escolhida a beta-glicosidase da lagarta *Spodoptera frugiperda* (Sfβgli).

Beta-glicosidases (EC 3.2.1.21) são enzimas que catalisam a hidrólise de ligações glicosídicas de aril e alquil-β-glicosídeos ou ainda de oligocelodextrinas, agindo no resíduo β-D-glucosídico não redutor e liberando D-glicose (Cairns e Esen, 2010; Lombard *et al.*, 2014). Elas estão presentes em diversos organismos, desempenhando diversas funções (Singh *et al.*, 2016). No entanto, apesar dessa variedade de funções, as beta-glicosidases pertencentes a GH1 são bastante conservadas, exibindo enovelamento do tipo barril (β/α)₈, mecanismo catalítico do tipo retentor da configuração do carbono anomérico e os mesmos resíduos catalíticos, dois ácidos glutâmicos que desempenham papel de nucleófilo e de ácido/base (Marana, 2006). No caso da Sfβgli, o resíduo que desempenha papel de ácido/base é o E187, e o resíduo que desempenha papel de nucleófilo catalítico é o E399, conforme indicado na Figura 5, contendo a estrutura primária e secundária da enzima.

Figura 5. Estrutura primária e terciária da beta-glicosidase de *Spodoptera frugiperda* (Sfβgli). Em A, a estrutura primária. Em B, a estrutura terciária. Em C, a estrutura com ênfase no sítio ativo da enzima. Os resíduos marcados em amarelo participam da catálise ou modulam o pK_a dos resíduos catalíticos. Os resíduos em vermelho interagem com a porção glicone do substrato. Os resíduos azuis interagem com a posição do aglicone do substrato. Em A, os resíduos catalíticos estão sublinhados. Figura extraída de Tamaki *et al.* (2016).

Além destes dois resíduos, a figura destaca outros resíduos que são conhecidamente importantes na catálise da enzima, modulando o pK_a dos resíduos

catalíticos, atuando na estabilização do estado de transição da catálise e na ligação do substrato, tanto na região que liga o glicone quanto na região que liga ao aglicone.

1.1.8 Relacionando centralidade com propriedades da Sfβgli

Neste contexto, recordando que o objetivo geral desta tese é avaliar experimentalmente a correlação entre a centralidade de resíduos com a sua importância nas propriedades de uma proteína, nesta etapa do trabalho o objetivo é tomar como modelo Sfβgli e produzir mutantes com substituições em seus resíduos centrais e também na vizinhança de um resíduo central. Após a produção e purificação das Sfβgli mutantes, seu enovelamento será analisado, indicando se podem ser usadas para caracterização funcional e estabelecimento de correlações entre centralidade e propriedades proteicas. Como critério de centralidade foi escolhido o parâmetro $C_{\Delta Lp}$, pois trata-se de um parâmetro de centralidade que relaciona resíduos externos ao sítio ativo com uma propriedade funcional de proteínas (Del_Sol *et al.*, 2006).

1.2 Materiais e Métodos

1.2.1 Produção de uma Rede de Estrutura de Proteínas para Sfβgli

A Rede de Estrutura de Sfβgli foi previamente construída e descrita na tese de C.M. Ikegami (2013). Esta presente tese desenvolve e aprofunda o estudo das Reses de Estrutura de Proteínas usando informações iniciais geradas naquele trabalho. Assim, considerou-se que a metodologia usada naquela tese deveria ser aqui mencionada para auxiliar o leitor.

Inicialmente, a estrutura tridimensional da Sfβgli foi modelada por homologia no servidor *Swiss Model* utilizando como molde a mirosinase de *Brevicoryne brassicae* (1WCG), de acordo com o descrito em Ikegami (2013). Após isso, o modelo de homologia (O61594) foi submetido a uma dinâmica molecular por 30 ns com campo de força OPLS-AA, com o apoio do prof. Dr. Guilherme Menegon Arantes (IQUSP). Depois dessa simulação de dinâmica molecular, o modelo médio da Sfβgli estável foi utilizado para a produção da Rede de Estrutura Proteica. Para isso, o mapa de contatos deste modelo médio de Sfβgli foi construído assumindo que dois resíduos estão conectados se cada um deles tivesse ao menos um átomo a 5Å de distância do outro, segundo sugerido por Greene e Higman (2003). Em seguida, os resíduos foram considerados como nós. Para esta rede foi determinado o caminho mínimo médio L_p empregando um programa produzido pelo Prof. Dr. Guilherme Menegon Arantes que se baseia no algoritmo de Floyd-Warshall. Para cálculo da centralidade de cada nó (resíduo), o parâmetro $C_{\Delta L_p}$ foi estimado segundo definido no item 1.1.5 (Del_Sol *et al.*, 2006). Os resíduos cuja remoção resultou em mudança do parâmetro L_p maior do que 2,9 desvios padrão foram considerados como centrais.

1.2.2 Linhagens de bactérias utilizadas no projeto

Durante o projeto, foram utilizados quatro tipos de linhagens de bactérias. Para o experimento de mutagênese sítio-dirigida, foi utilizado à linhagem *XL10 gold* segundo indicação do fabricante do *kit* de mutagênese (item 1.2.3). Para amplificação de plasmídeos, foi utilizada a linhagem *XL1 Blue*, que é mais apropriada para essa utilização

por ser deficiente na produção da endonuclease (*endA*) e em promover recombinações (deficiência em *recA*), promovendo a estabilidade do inserto.

1.2.3 Construções pET46_Sfβgli selvagem e mutantes de resíduos centrais

Para o início dos experimentos, foi utilizado uma construção previamente produzida por Ikegami (2013) contendo o inserto que codifica a Sfβgli selvagem em um vetor pET46 Ek/LIC (Novagen®, catálogo 71335-3). Desta forma, a construção possui promotor T7/lac, uma sequência que confere resistência a ampicilina, além de inserir uma cauda de histidina (His₆) na posição amino-terminal da proteína recombinante, o que permite a sua purificação com resina de níquel (Ni-NTA) (Figura 6). Além disso, foram utilizadas as construções pET46_Sfβgli D84A, R97A, L100A, T245A, S247A, N249A, F251A, F334A, L350A, S358A, K366A e Y420A previamente produzidas por (Ikegami, 2013). As demais construções foram obtidas segundo o item 1.2.3.

Figura 6. Representação da construção pET46_Sfβgli

1.2.4 Construções dos vetores contendo Sfβgli mutantes

Para execução do projeto foram feitas construções que codificam as Sfβgli mutantes D260A, E261A, M262A, A263F, A264F, E265A, A276F, G284F, V319F, R320A, L335A, V336F, S337F, L389A, Y390A e N391A. Para isso utilizou-se o kit *QuikChange™ Site Directed Mutagenesis* (Stratagene), seguindo-se as instruções do fabricante. Resumidamente, este procedimento empregou uma reação *polymerase chain reaction* (PCR) da construção inteira, tendo como molde a construção pET46_Sfβgli selvagem e como iniciadores mutagênicos as fitas de DNA indicadas na tabela 1, sintetizadas pela Integrated DNA Technologies (IDT). Após isso, o produto de PCR foi incubado com a enzima de restrição *DpnI* para digestão das fitas parentais, que são metiladas. O produto dessa reação é, então, utilizado para a transformação de bactérias competentes XL10 gold (segundo o item 1.3.5) para aumentar a quantidade de cópias.

Tabela 1. Iniciadores mutagênicos utilizados nos experimentos de mutagênese sítio-dirigida para produção das SfbgIi mutantes.

Mutação	Sequência do Iniciador
D260A	F 5' C ACG CCC ACT CCT GAG GCT GAA ATG GCT GCT GAA C 3' R 3' G TGC GGG TGA GGA CTC CGA CTT TAC CGA CGA CTT G 5'
E261A	F 5' G CCC ACT CCT GAG GAT GCA ATG GCT GCT GAA CTT AG 3' R 3' C GGG TGA GGA CTC CTA CGT TAC CGA CGA CTT GAA TC 5'
M262A	F 5' CC ACT CCT GAG GAT GAA GCG GCT GCT GAA CTT AGG AG 3' R 3' GG TGA GGA CTC CTA CTT CGC CGA CGA CTT GAA TCC TC 5'
A263F	F 5' C ACT CCT GAG GAT GAA ATG TTT GCT GAA CTT AGG AGA CAG 3' R 3' G TGA GGA CTC CTA CTT TAC AAA CGA CTT GAA TCC TCT GTC 5'
A264F	F 5' CCT GAG GAT GAA ATG GCT TTT GAA CTT AGG AGA CAG GGT G 3' R 3' GGA CTC CTA CTT TAC CGA AAA CTT GAA TCC TCT GTC CCA C 5'
E265A	F 5' GAG GAT GAA ATC GCT GCT GCA CTT AGG AGA CAG GGT GAG 3' R 3' CTC CTA CTT TAG CGA CGA CGT GAA TCC TCT GTC CCA CTC 5'
A276F	F 5' G GGT GAG TGG GGA ATC TAC TTT CAT CCT ATC TTC TCA GC 3' R 3' C CCA CTC ACC CCT TAG ATG AAA GTA GGA TAG AAG AGT CG 5'
G284F	F 5' CAT CCT ATC TTC TCA GCT GAA TTT GGT TTC CCC AAG GAG 3' R 3' GTA GGA TAG AAG AGT CGA CTT AAA CCA AAG GGG TTC CTC 5'
V319F	F 5' GAA GAA GAG AAG GCT TTC TTC AGA GGA ACT TCA GAT TTG 3' R 3' CTT CTT CTC TTC CGA AAG AAG TCT CCT TGA AGT CTA AAC 5'
R320A	F 5' GAA GAG AAG GCT TTC GTC GCA GGA ACT TCA GAT TTG ATC GG 3' R 3' CTT CTC TTC CGA AAG CAG CGT CCT TGA AGT CTA AAC TAG CC 5'
L335A	F 5' G AAC CAC TAC ACA GCA TCC GCG GTA TCG GCG ACT GAA CGC 3' R 3' C TTG GTG ATG TGT CGT AGG CGC CAT AGC CGC TGA CTT GCG 5'
V336F	F 5' CAC TAC ACA GCA TCC CTG TTT TCG GCG ACT GAA CGC 3' R 3' GTG ATG TGT CGT AGG GAC AAA AGC CGC TGA CTT GCG 5'
S337F	F 5' C TAC ACA GCA TCC CTG GTA TTT GCG ACT GAA CGC AAA GGT C 3' R 3' G ATG TGT CGT AGG GAC CAT AAA CGC TGA CTT GCG TTT CCA G 5'
L389A	F 5' G TTG ACC CAC CTG AAC AAC GCA TAC AAC AAA CCC GTC TTC 3' R 3' C AAC TGG GTG GAC TTG TTG CGT ATG TTG TTT GGG CAG AAG 5'
Y390A	F 5' G ACC CAC CTG AAC AAC TTA GCC AAC AAA CCC GTC TTC TAC 3' R 3' C TGG GTG GAC TTG TTG AAT CGG TTG TTT GGG CAG AAG ATG 5'
N391A	F 5' CC CAC CTG AAC AAC TTA TAC GCC AAA CCC GTC TTC TAC ATC 3' R 3' GG GTG GAC TTG TTG AAT ATG CGG TTT GGG CAG AAG ATG TAG 5'

1.2.5 Preparação de bactérias competentes

Para preparação de bactérias competentes uma alíquota de bactérias da linhagem apropriada para cada uso foi multiplicada através de crescimento em meio LB contendo tetraciclina até o cultivo atingir uma densidade óptica (DO_{600nm}) de aproximadamente 0,5. Após isso, os cultivos foram centrifugados e o meio de cultura descartado. As células foram ressuspensas em solução de $CaCl_2$ 100 $mmol \cdot L^{-1}$ gelado na proporção de 20 mL de solução para quantidade de células correspondente a 50 mL de cultivo. A mistura foi mantida em gelo durante 1 hora. Os tubos foram centrifugados e o sobrenadante descartado. As células foram ressuspensas em 1,0 mL de solução de $CaCl_2$ 100 $mmol \cdot L^{-1}$ gelado contendo glicerol 20% (v:v). Para armazenar as células, estas foram congeladas em alíquotas de 300 μL no caso da *XL10 gold* e 100 μL no caso das demais linhagens, e foram mantidas a $-80^\circ C$.

1.2.6 Transformação de bactérias *XL10 gold* por choque térmico

Para a transformação, as células *XL10 gold* competentes (preparadas segundo o item 1.2.5) foram misturadas a 6,0 μL de β -mercaptoetanol. Então, foram adicionados 10 μL do produto de reação do *kit* de mutagênese (item 1.2.4) à alíquota de bactérias competentes (item 1.2.5). Esta mistura foi incubada em gelo por 30 minutos, a $42^\circ C$ por 30 s, e após isso, em gelo por 2 min. Em seguida, foram adicionados 3,0 mL de meio LB suplementado com glicose 0,4% (m:v), $MgCl_2$ 12,5 $mmol \cdot L^{-1}$ e $MgSO_4$ 12,5 $mmol \cdot L^{-1}$ aquecido a $42^\circ C$. As células foram, então, incubadas a $37^\circ C$ e agitação de 250 rpm durante 1 hora. Ao final, as células foram transferidas para placas de LB ágar, contendo

carbenicilina $100 \text{ mg}\cdot\text{L}^{-1}$. As placas foram incubadas em estufa a 37°C durante aproximadamente 16 h.

1.2.7 Transformação de bactérias BL21 gold(DE3), XL1 Blue e Nova Blue (DE3)

Para transformar as bactérias da linhagem *BL21 gold (DE3)*, *XL1 Blue* e *Nova Blue (DE3)* foram adicionados $2,0 \text{ }\mu\text{L}$ de preparação plasmidial contendo a construção pET46_Sfβgli selvagem (ou pET46_Sfβgli mutantes) em uma alíquota de bactérias competentes. As células, em seguida, foram mantidas em gelo durante 30 min. Após isso, as células foram transferidas para um banho a 42°C durante 90 s, e em seguida colocadas novamente em gelo por 2 min. Então, $500 \text{ }\mu\text{L}$ de meio LB foram adicionados às células. Estas foram mantidas a 37°C com agitação de 200 rpm durante 1 h. O conteúdo dos tubos foi transferido para meio sólido LB-ágar contendo carbenicilina $100 \text{ mg}\cdot\text{L}^{-1}$. As placas foram mantidas em estufa a 37°C durante 16 h, para o crescimento de colônias.

1.2.8 Preparação de DNA plasmidial em maior escala

Colônias de *XL10 gold* e *XL1Blue* após a transformação com vetor de interesse foram retiradas com palito estéril e transferidas para 10 mL de meio LB contendo carbenicilina $100 \text{ mg}\cdot\text{L}^{-1}$. Os cultivos foram mantidos a 37°C com agitação durante 16 h, e posteriormente centrifugados, sendo que o sobrenadante foi descartado e os plasmídeos foram purificados do sedimento utilizando o *kit* “Wizard Plus SV Minipreps DNA Purification System”, da Promega Corporation.

1.2.9 Análise dos vetores que codificam Sfβgli mutantes

A presença do inserto que codifica Sfβgli nos vetores purificados foi confirmada através da amplificação em reação de PCR utilizando iniciadores 5N (5' GACGACGACAAG ATCCAGCAGCGCC 3') e 3C (5' GAGGAGAAGCCCGGTTACGGGGCG 3') que pareiam nas extremidades do inserto. As reações utilizaram o reagente *PCR mastermix* (Fermentas).

Ao final foi feita eletroforese em gel de agarose com o produto de PCR para confirmar a presença da banda correspondente ao inserto da Sfβgli. Para isso foi utilizado gel de agarose 1,0% (m:v) em tampão Tris-acetato 40 mmol·L⁻¹ contendo EDTA 1,0 mmol·L⁻¹ (tampão TAE). As amostras foram aplicadas com adição de tampão de amostra, na proporção de 1,0 µL de *Blue Green Loading Dye* (LGC Biotecnologia) para cada 5,0 µL de amostra. Esse tampão de amostra contém substância que emite fluorescência e interage com DNA, permitindo a visualização de fitas de DNA.

A eletroforese foi feita a 80 V durante 1 hora. Ao final o gel teve sua imagem capturada através de um foto-documentador (BioRad). Após confirmar a presença do inserto nas amostras, elas foram submetidas a um sequenciamento no Centro de Sequenciamento de DNA do IQ-USP para confirmar a presença da mutação de interesse e ausência de mutações espúrias.

1.2.10 Expressão de Sfβgli selvagem e mutantes

Para produção de Sfβgli recombinante, colônias de bactérias *BL21 gold(DE3)* ou *Nova Blue (DE3)* previamente transformadas segundo o item 1.2.7 foram retiradas com palitos estéreis e transferidas para dois tubos contendo 10 mL de meio LB e carbenicilina

100 mg·L⁻¹, onde foram incubadas durante aproximadamente 18 h a 37 °C, com agitação de 200 rpm.

Então, esse cultivo foi inoculado em *erlenmeyers* contendo 500 mL de meio LB e carbenicilina 100 mg·L⁻¹. Os cultivos foram incubados a 20 °C com agitação de 200 rpm. O crescimento das bactérias foi monitorado através da turbidez, medida pela absorbância a 600 nm (Abs₆₀₀) do cultivo. Após atingir Abs₆₀₀ entre 0,5 e 1,0 adicionou-se isopropil β-D-1-tiogalactopiranosideo (IPTG) aos cultivos para indução, de forma que a concentração final de 1,0 mmol·L⁻¹ fosse atingida. Os cultivos foram mantidos a 20 °C com agitação de 200 rpm por 20 h. Após isso, foram centrifugados, e o sobrenadante contendo o meio de cultivo foi descartado. O sedimento contendo as células foi coletado e congelado a -80 °C para posterior tratamento. Uma alíquota do material antes e depois da indução foi analisada por SDS-PAGE, segundo o item 1.2.12.

1.2.11 Lise de bactérias e purificação de Sfβgli selvagem e mutantes

As bactérias induzidas (item 1.2.10) foram lisadas ressuspendendo-as em tampão fosfato de sódio 50 mmol·L⁻¹pH 8,0 contendo cloreto de sódio 300 mmol·L⁻¹ e imidazol 20 mmol·L⁻¹, na proporção de 1,0 mL de tampão de lise para cada 50 mL de cultivo. Após isso, as células foram lisadas através de ultrassom empregando cinco pulsos de 10 s cada a 10 % de potência (Branson Sonifier 250, Soni-tech), com intervalos em gelo de pelo menos 1 min. Então os restos de célula foram removidos centrifugando a suspensão a 9.000 rpm por 30 min a 4 °C. O sobrenadante, contendo a enzima, foi coletado, e 5,0 µL deste foram misturados com 15 µL de tampão de amostra de SDS-PAGE, para posterior análise, segundo

o item 1.2.12. O sedimento foi ressuspendido em 1,0 mL de tampão de lise e uma alíquota de 5,0 μL desse material e do material solúvel (sobrenadante recolhido) foram misturadas com 15 μL de tampão de amostra de SDS-PAGE para serem analisadas, segundo o item 1.2.12.

A Sf β gli presente no lisado foi purificada através de cromatografia de afinidade com a resina Ni-NTA agarose (Qiagen). Para isso, a resina foi incubada com o lisado na proporção de 200 μL de resina para 1,0 mL de lisado, durante 16 h a 4 °C com agitação de um rotor de torque. Após isso, a suspensão foi centrifugada a 5.000 rpm por 1 min e o sobrenadante, chamado de *Flow-Through* (FT), foi analisado por SDS-PAGE aplicando-se 5,0 μL em 15 μL de tampão de amostra, segundo o item 1.2.12. A resina foi lavada com tampão fosfato de sódio 50 $\text{mmol}\cdot\text{L}^{-1}$ pH 8,0 contendo cloreto de sódio 300 $\text{mmol}\cdot\text{L}^{-1}$ e imidazol 20 $\text{mmol}\cdot\text{L}^{-1}$. A primeira lavagem, após ser separada da resina por centrifugação, foi analisada por SDS-PAGE, adicionando-se 5,0 μL da lavagem e 15 μL de tampão de amostra. Foram executadas 10 lavagens. Então, a enzima foi eluída da resina adicionando-se 120 μL (para cada 200 μL de resina) de tampão fosfato de sódio 50 $\text{mmol}\cdot\text{L}^{-1}$ pH 7,0 contendo cloreto de sódio 300 $\text{mmol}\cdot\text{L}^{-1}$ e imidazol 300 $\text{mmol}\cdot\text{L}^{-1}$, e incubando por 15 min a 4 °C com agitação. Ao final, a suspensão foi centrifugada e o sobrenadante, contendo a Sf β gli, foi coletado. Foram executadas três eluições com a mesma resina. O conjunto das eluições foi analisado por SDS-PAGE. A resina após a purificação também foi analisada por SDS-PAGE. Para realizar os ensaios de dicroísmo circular, fluorescência dos triptofanos e DSF, o tampão da amostra foi trocado por tampão fosfato de potássio 20 $\text{mmol}\cdot\text{L}^{-1}$ pH 6,0 empregando coluna de dessalinização *Hi-TrapTMDesalting* (GE Healthcare), seguindo-se instruções do

manual. Já para os ensaios de caracterização cinética das enzimas, o tampão foi trocado por citrato-fosfato 100 mmol^{-1} pH 6,0, também utilizando a coluna de dessalinização.

1.2.12 SDS-PAGE

Para avaliar a indução e a pureza do material purificado, foram feitas eletroforeses em condições desnaturantes (SDS-PAGE). Para isso, foram preparados géis de poliacrilamida 10% e SDS 0,1% segundo Laemmli (1970).

Antes de aplicar as alíquotas de proteína no gel, estas foram desnaturadas, adicionando tampão de amostra (TRIS-HCl $60 \text{ mmol} \cdot \text{L}^{-1}$ pH 6,8, SDS 2,5% (m:v), β -mercaptoetanol $0,36 \text{ mmol} \cdot \text{L}^{-1}$, EDTA $0,5 \text{ mmol} \cdot \text{L}^{-1}$, glicerol 10% (v:v) e azul de bromofenol 0,005% (m:v)) em volume igual ou superior ao da alíquota da amostra, e aquecendo-se esta mistura a 95°C por 5 min. Os géis foram colocados em uma cuba de eletroforese, que foi preenchida com solução-tampão de corrida (TRIS $25 \text{ mmol} \cdot \text{L}^{-1}$, glicina $0,25 \text{ mol} \cdot \text{L}^{-1}$ e SDS 0,1%). O sistema foi ligado em uma fonte a 200 V, durante aproximadamente 40 min.

Para a visualização das proteínas, os géis foram corados com Comassie Blue R-250 0,1% (m:v) preparado em uma solução de etanol 40% (v:v) e ácido acético 10% (v:v), imergindo os géis nesta solução durante 30 min. Em seguida os géis foram descorados com solução de etanol 40% (v:v) e ácido acético 10% (v:v). Ao final, os géis foram mantidos em água, e fotografados através de foto-documentador.

1.2.13 Determinação da concentração de Sfβgli

Foram feitas medidas de concentração da Sfβgli através de sua absorção de radiação na região do ultravioleta (UV) a 280 nm. Primeiramente, a absortividade molar (ϵ) da Sfβgli desnaturada foi calculada através do aplicativo *online ProtParam tool* (ExPASy). Então, alíquotas do material contendo a Sfβgli foram colocadas em excesso (proporção de no mínimo 1:100) de solução desnaturante (cloreto de guanidina $6,0 \text{ mol}\cdot\text{L}^{-1}$ preparada em solução-tampão de fosfato $20 \text{ mmol}\cdot\text{L}^{-1}$ pH 6,0). Em seguida, determinou-se a absorbância deste material em um espectrofotômetro de UV-visível (Ultrospec 2100 pro) em um comprimento de onda de 280 nm utilizando-se uma cubeta de quartzo de caminho óptico de 1,0 cm. Através do ϵ_{280} ($120210 \text{ L}\cdot\text{mol}^{-1}\cdot\text{cm}^{-1}$) a concentração foi estimada pela lei de Beer-Lambert, expressa pela equação 6, na qual Abs é a absorbância, ϵ é a absortividade molar, b é o caminho óptico em cm, e c é a concentração do material na cubeta.

Equação 6

$$Abs = \epsilon \cdot b \cdot c$$

1.2.14 Coleta de espectros de fluorescência de triptofano

Para avaliar se a Sfβgli purificada se encontrava enovelada, foram coletados espectros de fluorescência de triptofano. Os espectros de fluorescência foram coletados em um equipamento *Fluorescence Spectrophotometer f-4500* (Hitachi) com excitação de 295 nm e emissão de 310 a 450 nm utilizando a velocidade de varredura de $60 \text{ nm}\cdot\text{min}^{-1}$ a 25°C . Foi empregada cubeta de quartzo com quatro faces transparentes e caminho óptico de absorção de 1,0 cm.

1.2.15 Apagamento da fluorescência dos triptofanos

Para avaliar a exposição dos triptofanos da Sfβgli selvagem e mutantes, foi executado ensaio de apagamento de fluorescência dos triptofanos por acrilamida. Para isso, uma solução estoque de acrilamida $10,0 \text{ mol}\cdot\text{L}^{-1}$ foi adicionada a $500 \mu\text{L}$ da amostra, sendo que foram adicionados $25 \mu\text{L}$ de forma fracionada (com adições de 2, 4 e $5 \mu\text{L}$). Após cada adição, a amostra foi termostatizada a 25°C por 5 min, e a fluorescência com excitação de 295 nm e emissão de 341 nm foi coletada. Para a análise dos dados foram plotados gráficos da razão entre fluorescência inicial e após a adição de supressor *versus* a concentração de supressor. O parâmetro K_{sv} é a inclinação da reta obtida, conforme a equação de Stern-Volmer (equação 7) (Eftink e Ghiron, 1981).

$$\frac{I_f^0}{I_f} = 1 + K_{sv} \cdot [Q] \quad \text{Equação 7}$$

Onde I_f^0 é a intensidade de fluorescência inicial, I_f é a intensidade de fluorescência após a adição de supressor, K_{sv} é o parâmetro de Stern-Volmer, e $[Q]$ é a concentração de supressor.

1.2.16 Coleta de espectros de Dicroísmo Circular

Para avaliar a estrutura secundária da Sfβgli, foram coletados espectros de dicroísmo circular (CD) das amostras purificadas. Para isso, as amostras foram diluídas para uma concentração final de $0,2 \text{ g}\cdot\text{L}^{-1}$. O espectro foi coletado em equipamento Espectropolarímetro J-720 (JASCO, Inc) na faixa de 260 a 190 nm com velocidade de varredura de $20 \text{ nm}/\text{min}$ e acumulação de 8 espectros para aumento do sinal/ruído. Para o

cálculo da elipsidade molar, foi utilizada a equação 8, onde $[\theta]$ é a elipsidade molar residual ($\text{mDeg}\cdot\text{cm}^2\cdot\text{dmol}^{-1}$), θ é a elipsidade em mDeg , PM é o peso molecular da proteína em Da, l é o caminho óptico da cubeta utilizada em cm, c é a concentração da proteína em $\text{mg}\cdot\text{mL}^{-1}$ e NR é o número de resíduos de aminoácido da proteína (Kelly *et al.*, 2005). O resultado foi analisado através do programa K2d3 (Louis-Jeune *et al.*, 2012).

$$[\theta] = \frac{\theta \cdot PM}{10 \cdot l \cdot c \cdot (NR - 1)} \quad \text{Equação 8}$$

1.3 Resultados e Discussão

1.3.1 Planejamento das mutações em Sfβgli para teste da correlação centralidade e função proteica

Del Sol e colaboradores (2004) descreveram uma correlação entre centralidade de resíduos externos ao sítio com uma função proteica. Por conta disso, este parâmetro de centralidade ($C_{\Delta Lp}$) foi escolhido nesta abordagem experimental que emprega uma enzima como modelo, uma vez que nosso grupo se interessa pela influência do entorno nas propriedades dos sítios ativos. Naturalmente há aspectos práticos na escolha deste parâmetro de centralidade, uma vez que a construção da rede de estrutura de Sfβgli e os cálculos de centralidade por $C_{\Delta Lp}$ já estavam disponíveis no nosso grupo.

Deste modo, inicialmente a estrutura de Sfβgli obtida a partir de uma modelagem por homologia elaborada no servidor *Swiss Model* (Modelo: O61594) foi submetida a uma dinâmica molecular e então transformada em rede (Ikegami, 2013; Souza *et al.*, 2016).

Resumidamente, a estrutura média obtida da dinâmica molecular foi transformada em rede considerando-se que dois resíduos fazem contato se tiverem átomos a uma distância de no máximo 5 Å um do outro, de acordo com o proposto por Greene e Higman (2003) (Figura 7). Com base nessa rede, a centralidade dos resíduos foi calculada computacionalmente avaliando o efeito que a deleção de cada um dos resíduos tinha no caminho mínimo médio da rede proteica (L_p), ou seja, no número mínimo de conexões médio necessário para ir de um resíduo até outro qualquer da rede (Del_Sol *et al.*, 2006). Os resíduos cuja deleção resultou em um aumento de L_p em pelo menos 2,9 desvios padrão do aumento médio foram considerados centrais da rede. Desta forma, 11 dos 509 resíduos da Sfβgli foram considerados centrais, sendo eles R97, T245, S247, N249, F251, Y331, F334, S358, K366, E399 e Y420 (Ikegami, 2013; Souza *et al.*, 2016). Conforme as citações acima, a construção da rede estrutural de Sfβgli e a identificação dos resíduos centrais foi previamente reportada na tese de C.M. Ikegami (Ikegami, 2013).

Conhecidos os resíduos centrais foi considerado que para avaliar a correlação entre estes e as propriedades da Sfβgli, é necessário perturbar estes resíduos centrais e caracterizar os efeitos sobre propriedades globais da proteína. Esta perturbação pode ser feita de duas formas diferentes, modificando o próprio resíduo central ou modificando a vizinhança deste resíduo. Nesta tese, utilizamos estas duas estratégias.

Figura 7. Rede de Estrutura Proteica da Sfβgli. Em A, o modelo de homologia representado como uma rede. As esferas pretas representam cada um dos resíduos na posição do carbono-alfa. Em B, representação circular da rede. Os resíduos são representados por pontos pretos distribuídos ao redor do círculo no sentido horário, onde o traço preto representa o resíduo 1. Traços verdes representam as interações. Extraído de Souza *et al.* (2016).

Na primeira, mutações foram direcionadas aos próprios resíduos centrais de forma a substituí-los por alanina. Como alanina possui uma cadeia lateral curta e pouco volumosa, ela faz poucos contatos com a vizinhança e, portanto, modificar um resíduo central para alanina significa diminuir a conectividade deste resíduo, o que pode em tese perturbar sua participação como resíduo central e assim a rede como um todo. Desta forma, para produção de Sfβgli mutantes carregando as substituições diretas em resíduos centrais T245A, S247A, N249A, F251A, F334A, S358A, K366A e Y420A foram empregados vetores previamente construídos por Ikegami (2013). As mutações nos resíduos R97, Y331 e E399 não foram feitas porque sendo resíduos do sítio ativo da Sfβgli (Marana *et al.*, 2003), não seria possível creditar os eventuais efeitos destas mutações à centralidade ou participação direta na catálise. Além disso, também foram empregados vetores que codificam Sfβgli

com mutações em resíduos não-centrais (L100A, R320A e L350A; produzido por Ikegami (2013); e L389A, Y390A e N391A; produzidos nesta tese) para compor um grupo controle. Na figura 8 A estão indicados todos os resíduos mutados segundo esta estratégia e o efeito que a substituição por alanina causa na conectividade com a vizinhança.

Além dessa estratégia, também foram feitas mutações ao redor do resíduo central F251, para perturbar o ambiente ao qual está inserido. Essas mutações foram dirigidas apenas a vizinhos do F251 que não fossem centrais. Isto garante que eventuais efeitos avaliados sejam provavelmente resultantes da perturbação do resíduo central e não da centralidade do próprio resíduo modificado. Ainda essas mutações foram planejadas de forma a introduzir volume extra em resíduos de cadeia lateral pouco volumosas e vice-versa, na tentativa de afetar o resíduo central F251, modificando seu posicionamento. Com base neste raciocínio foram então planejadas as mutações D260A, E261A, M262F, A263F, A264F, E265A, que ficam em uma alfa-hélice vizinha ao F251, e as mutações L335A, V336F e S337A, que ficam em uma fita beta, também vizinha ao F251. Na figura 8B estão representados os resíduos vizinhos do F251 mutados durante este trabalho, além de um esquema ilustrando um exemplo de perturbação provocada no F251 por uma mutação em sua vizinhança. Nota-se no caso que a substituição de V336 por F gera potencialmente um impedimento estérico com F251.

Figura 8. Estratégias de perturbações dos resíduos centrais da Sfβgli. Em A estão representados na estrutura os 8 resíduos centrais (vermelhos) e os 5 resíduos não-centrais (azuis) a receberem mutação direta, além do efeito provocado pela substituição de um resíduo central por alanina. Em B, a vizinhança do resíduo central F251, indicando os resíduos mutados (cinza), e na parte inferior uma representação do efeito provocado por uma das mutações na vizinhança do resíduo F251.

1.3.2 Construção das Sfβgli mutantes

As mutações sítio-dirigidas que codificam as substituições D260A, E261A, M262A, A263F, A264F, E265A, R320A, L335A, V336F, S337F, F389A, Y390A e N391A foram introduzidas na construção que codifica Sfβgli selvagem seguindo-se as condições propostas pelo *kit Quik Site Directed Mutagenesis*, mas com algumas adaptações, segundo o item 1.2.4.

Para verificar se as colônias obtidas ao final do procedimento de mutagênese continham, de fato, o vetor contendo inserto, foram feitas extrações plasmidiais seguidas de reações de PCR para amplificar o segmento de DNA que codifica a Sfβgli. Os produtos amplificados nestas reações foram submetidos a uma eletroforese em gel de agarose e as fitas de DNA correspondentes ao inserto da Sfβgli (~1,5 kb) foram visualizadas conforme pode ser observado na figura 9 (o procedimento detalhado está descrito no item 1.2.9).

Figura 9. Eletroforese em gel de agarose do produto de reações de PCR. Nas raias temos, respectivamente, padrão, o produto de PCR da Sfβgli selvagem e dos vetores codificando os mutantes D260A, E261A, M262A, A263F, A264F, E265A, R320A, L335A, V336F, S337F, F389A, Y390A e N391A.

Na figura 9 é possível observar que todos os vetores contêm um inserto com tamanho compatível para codificar Sfβgli. A seguir para verificar se as mutações foram inseridas corretamente, os insertos foram sequenciados. A sequência codificadora de Sfβgli selvagem, determinada experimentalmente, está representada na figura 10. Na figura 11, estão trechos do sequenciamento dos insertos que codificam as Sfβgli mutantes. Apesar de sequenciadas por completo, na figura está ilustrado somente o trecho contendo a mutação de interesse. Conforme ilustrado, as mutações foram inseridas corretamente em todos os casos. Além disso, não ocorreram mutações espúrias em outros pontos do inserto.

5	ATG	GCA	CAT	CAC	CAC	CAC	CAT	CAC	GTG	GAT	GAC	GAC	GAC	AAG	ATC	20
	M	A	H	H	H	H	H	H	V	D	D	D	D	K	I	
21	CAG	CAG	CGC	CGC	TTC	CCC	GAT	GAC	TTC	CTC	TTC	GGA	ACC	GCT	ACA	35
	Q	Q	R	R	F	P	D	D	F	L	F	G	T	A	T	
36	GCC	TCC	TAC	CAG	ATC	GAA	GGT	GCT	TGG	GAT	GAA	GAT	GGT	AAG	GGT	50
	A	S	Y	Q	I	E	G	A	W	D	E	D	G	K	G	
51	GAA	AAC	ATC	TGG	GAC	TAC	ATG	GTC	CAT	AAC	ACC	CCA	GAA	GTC	ATC	65
	E	N	I	W	D	Y	M	V	H	N	T	P	E	V	I	
66	AGG	GAT	CTA	TCC	AAC	GGT	GAC	ATC	GCT	GCT	GAC	TCC	TAC	CAC	AAC	80
	R	D	L	S	N	G	D	I	A	A	D	S	Y	H	N	
81	TAC	AAG	CGT	GAT	GTC	GAG	ATG	ATG	AGG	GAG	TTA	GGC	CTG	GAC	GCT	95
	Y	K	R	D	V	E	M	M	R	E	L	G	L	D	A	
96	TAC	AGG	TTC	TCC	CTC	TCC	TGG	GCT	CGG	ATC	CTG	CCC	ACT	GGC	ATG	110
	Y	R	F	S	L	S	W	A	R	I	L	P	T	G	M	
111	GCC	AAT	GAG	GTC	AAC	CCA	GCT	GGC	ATC	GCT	TTT	TAC	AAC	AAC	TAC	125
	A	N	E	V	N	P	A	G	I	A	F	Y	N	N	Y	
126	ATT	GAT	GAA	ATG	CTG	AAG	TAC	AAC	ATC	ACT	CCT	CTG	ATC	ACG	CTC	140
	I	D	E	M	L	K	Y	N	I	T	P	L	I	T	L	
141	TAC	CAT	TGG	GAT	CTT	CCT	CAA	AAG	CTG	CAG	GAG	TTG	GGT	GGT	TTC	155
	Y	H	W	D	L	P	Q	K	L	Q	E	L	G	G	F	
156	GCC	AAT	CCC	CTC	ATC	AGC	GAC	TGG	TTC	GAA	GAC	TAC	GCG	AGG	GTT	170
	A	N	P	L	I	S	D	W	F	E	D	Y	A	R	V	
171	GTC	TTC	GAG	AAC	TTC	GGA	GAC	AGA	GTC	AAG	ATG	TTC	ATC	ACT	TTC	185
	V	F	E	N	F	G	D	R	V	K	M	F	I	T	F	
186	AAC	GAG	CCT	AGA	GAG	ATT	TGC	TTT	GAG	GGT	TAT	GGA	TCT	GCT	ACC	200
	N	E	P	R	E	I	C	F	E	G	Y	G	S	A	T	
201	AAG	GCT	CCG	ATC	CTA	AAC	GCT	ACC	GCG	ATG	GGA	GCT	TAC	TTG	TGT	215
	K	A	P	I	L	N	A	T	A	M	G	A	Y	L	C	
216	GCC	AAG	AAC	TTG	GTG	ACT	GCT	CAT	GCC	AAG	GCT	TAC	TAT	TTG	TAT	230
	A	K	N	L	V	T	A	H	A	K	A	Y	Y	L	Y	
231	GAC	AGA	GAG	TTC	AGG	CCT	GTG	CAG	GGT	GGT	CAG	TGC	GGT	ATT	ACC	245
	D	R	E	F	R	P	V	Q	G	G	Q	C	G	I	T	
246	ATC	AGT	GTA	AAC	TGG	TTT	GGA	CCA	GCC	ACG	CCC	ACT	CCT	GAG	GAT	260
	I	S	V	N	W	F	G	P	A	T	P	T	P	E	D	
261	GAA	ATG	GCT	GCT	GAA	CTT	AGG	AGA	CAG	GGT	GAG	TGG	GGA	ATC	TAC	275
	E	M	A	A	E	L	R	R	Q	G	E	W	G	I	Y	

276	GCT	CAT	CCT	ATC	TTC	TCA	GCT	GAA	GGA	GGT	TTC	CCC	AAG	GAG	CTT	290
	A	H	P	I	F	S	A	E	G	G	F	P	K	E	L	
291	TCA	GAC	AAG	ATC	GCC	GAA	AAG	AGC	GCT	CAG	CAA	GGC	TAC	CCT	TGG	305
	S	D	K	I	A	E	K	S	A	Q	Q	G	Y	P	W	
306	TCT	CGT	CTG	CCA	GAA	TTC	ACT	GAA	GAA	GAG	AAG	GCT	TTC	GTC	AGA	320
	S	R	L	P	E	F	T	E	E	E	K	A	F	V	R	
321	GGA	ACT	TCA	GAT	TTG	ATC	GGA	GTG	AAC	CAC	TAC	ACA	GCA	TTC	CTG	335
	G	T	S	D	L	I	G	V	N	H	Y	T	A	F	L	
336	GTA	TCG	GCG	ACT	GAA	CGC	AAA	GGT	CCG	TAC	CCT	GTG	CCA	TCT	TTA	350
	V	S	A	T	E	R	K	G	P	Y	P	V	P	S	L	
351	CTG	GAT	GAT	GTG	GAC	ACT	GGC	TCC	TGG	GCT	GAT	GAT	AGC	TGG	CTC	365
	L	D	D	V	D	T	G	S	W	A	D	D	S	W	L	
366	AAG	TCT	GCT	TCA	GCT	TGG	CTT	ACA	CTT	GCA	CCA	AAC	AGC	ATC	CAC	380
	K	S	A	S	A	W	L	T	L	A	P	N	S	I	H	
381	ACC	GCG	TTG	ACC	CAC	CTG	AAC	AAC	TTA	TAC	AAC	AAA	CCC	GTC	TTC	395
	T	A	L	T	H	L	N	N	L	Y	N	K	P	V	F	
396	TAC	ATC	ACT	GAG	AAT	GGC	TGG	TCT	ACT	GAC	GAG	AGC	CGA	GAG	AAC	410
	Y	I	T	E	N	G	W	S	T	D	E	S	R	E	N	
411	AGT	CTC	ATT	GAT	GAT	GAC	AGG	ATC	CAG	TAC	TAC	AGG	GCT	TCC	ATG	425
	S	L	I	D	D	D	R	I	Q	Y	Y	R	A	S	M	
426	GAG	AGT	CTC	CTG	AAC	TGC	CTG	GAT	GAT	GGC	ATC	AAC	CTC	AAA	GGA	440
	E	S	L	L	N	C	L	D	D	G	I	N	L	K	G	
441	TAC	ATG	GCT	TGG	AGT	CTA	ATG	GAC	AAC	TTT	GAA	TGG	ATG	GAG	GGT	455
	Y	M	A	W	S	L	M	D	N	F	E	W	M	E	G	
456	TAT	ATT	GAG	CGT	TTC	GGC	CTT	TAC	GAA	GTA	GAT	TTC	TCG	GAC	CCC	470
	Y	I	E	R	F	G	L	Y	E	V	D	F	S	D	P	
471	GCG	CGC	ACG	CGC	ACC	CCC	AGG	AAG	GCA	GCG	TTC	GTG	TAC	AAA	CAC	485
	A	R	T	R	T	P	R	K	A	A	F	V	Y	K	H	
486	ATC	ATC	AAA	CAT	CGC	GTC	GTG	GAC	TAT	GAA	TAC	GAA	CCG	GAG	ACC	500
	I	I	K	H	R	V	V	D	Y	E	Y	E	P	E	T	
501	ATG	GTC	ATG	ACT	ATA	GAT	GAG	GGA	CAT	TGA						
	M	V	M	T	I	D	E	G	H	Stop						

Figura 10. Sequência de DNA completa que codifica a Sfβgli selvagem clonada no vetor pET46. Em preto está a sequência de nucleotídeos do inserto, determinada experimentalmente. As letras coloridas são os respectivos resíduos de aminoácidos, após a tradução.

Figura 11. Sequenciamento de DNA do inserto que codifica Sfβgli selvagem e mutantes. Ao fundo está o sequenciamento do segmento de DNA que codifica a enzima selvagem e nos retângulos em destaque, trechos do sequenciamento dos insertos mutantes contendo a substituição de interesse (grifada em vermelho).

Após o sequenciamento, foram obtidas maiores quantidades dos plasmídeos que codificam as Sfβgli selvagem e mutantes, amplificando-os em bactérias *XL1Blue* e purificando-os utilizando o *kit Wizard Miniprep*, segundo o item 1.2.8.

1.3.3 Expressão e purificação de Sfβgli selvagem e mutantes

A Sfβgli selvagem e todos os mutantes foram expressos como proteínas recombinantes utilizando as linhagens *BL21 gold(DE3)* e *Nova Blue(DE3)*, conforme o procedimento descrito no item 1.2.10. Para verificar se houve indução após adição de IPTG, foram feitos SDS-PAGE de uma alíquota de cultivo antes e depois da indução. Como pode ser observado, na maior parte dos casos houve expressão de grande quantidade das Sfβgli mutantes (Figura 12), exceto no caso dos mutantes N249A e F334A, que apresentaram bandas menos intensas.

Após isso, as bactérias induzidas foram lisadas e as Sfβgli selvagem e mutantes foram purificados pelo método de *batch* utilizando-se resina Ni-NTA, descrito no item 1.2.11. As amostras coletadas durante a purificação foram analisadas por SDS-PAGE (Figura 13). Como pode ser observado, a Sfβgli selvagem e a maior parte das mutantes foram purificadas satisfatoriamente. No entanto, não foi possível purificar os mutantes R320A e T245A, pois após a lise, permaneceram na fração insolúvel, sugerindo que durante a etapa de expressão estas proteínas mutantes ficaram retidas em corpos de inclusão. Destes, T245A, contém uma mutação em resíduo central e R320A contém mutação em um resíduo não-central. Além da expressão de T245A em *BL21 gold(DE3)*, também foi testada a expressão deste mutante em bactérias *Nova Blue (DE3)*, mas o processo também não foi satisfatório e a proteína continuou sendo acumulada

em corpos de inclusão. Embora este efeito possa estar relacionado à centralidade do resíduo, é incerta essa correlação, já que a mutação no resíduo não-central R320A teve o mesmo efeito.

Figura 12. Análise da produção das Sfβgli selvagem e mutantes recombinantes em *BL21 gold (DE3)* através de SDS-PAGE. As raias marcadas por NI são as amostras de culturas de bactérias antes da indução com IPTG. As raias I são as amostras de cultivo após a adição de IPTG. Na primeira linha estão amostras de culturas bacterianas que produzem a proteína selvagem e os mutantes controle (substituições em resíduos não-centrais), na segunda linha estão as mutações dirigidas diretamente aos resíduos centrais, e na terceira linha, as mutações na vizinhança do resíduo central F251.

De qualquer modo, tendo 13 das 15 proteínas mutantes sido purificadas satisfatoriamente, sendo 7 contendo mutações em resíduos centrais e 6 em resíduos não-centrais, consideramos a amostra suficiente para prosseguir nas análises, mesmo sem a purificação dos mutantes T245A e R320A (Figura 13).

Figura 13. Análise em SDS-PAGE da purificação da Sfβgli selvagem e mutantes utilizando resina de Ni-NTA Agarose. Extraído de Souza *et al.* (2016).

Após a purificação, o tampão das amostras de proteína foi removido e trocado por fosfato de potássio 10,0 mM pH 6,0, utilizando colunas de dessalinização, o que resultou em redução de suas concentrações devido à diluição. As concentrações finais (após a dessalinização) das amostras obtidas utilizando as linhagens *BL21 gold(DE3)* e *Nova Blue(DE3)* para produção de

cada um dos mutantes estão representadas na tabela 2. Como pode ser observado, a linhagem *BL21 gold(DE3)* produziu uma maior quantidade de proteínas, e por isso foi escolhida para produzir amostras para os ensaios de verificação do dobramento proteico, dicroísmo circular e fluorescência intrínseca de triptofanos. A linhagem *Nova Blue (DE3)* foi utilizada somente para a produção de amostras para caracterização enzimática (ver capítulo 3), evitando-se assim a contaminação das amostras de Sfβgli por β-galactosidase, a qual poderia ser eventualmente ativa sobre substratos da primeira e então resultar em parâmetros cinéticos incorretos.

1.3.4 Espectro de fluorescência de triptofanos da Sfβgli selvagem e mutantes

Para avaliar se as proteínas purificadas estavam enoveladas corretamente foram coletados espectros de fluorescência de triptofanos das Sfβgli mutantes e estes foram comparados com o espectro da enzima selvagem. Como pode ser observado na figura 14, nenhum dos espectros de emissão de fluorescência das enzimas mutantes apresentou diferenças significativas com relação à enzima selvagem, sugerindo que todos, tanto mutantes que perturbam os resíduos centrais quanto os controles estão enovelados.

Tabela 2. Concentração das Sfβgli selvagem e mutantes recombinantes após a purificação.

Sfβgli	Concentração (g·L ⁻¹)	
	Amostra produzida em BL21 gold (DE3)	Amostra produzida em Nova Blue (DE3)
Selvagem	0,96	0,52
T245A	-	-
S247A	0,31	*
N249A	0,077	*
F251A	0,18	*
F334A	0,87	*
S358A	0,79	*
K366A	0,90	*
Y420A	0,14	*
L100A	0,14	0,15
L350A	0,85	0,34
D260A	0,33	0,38
E261A	0,39	0,43
M262A	1,0	0,73
A263F	0,31	0,18
A264F	0,13	
E265A	0,55	0,41
L335A	1,4	0,55
V336F	0,24	
S337F	0,36	0,11
R320A	-	-
L389A	0,82	0,37
Y390A	0,10	
N391A	0,26	0,53

* Esses mutantes foram produzidos em Nova Blue (DE3) por Ikegami (2013).

Figura 14. Espectros de emissão de fluorescência de triptofanos da Sfβgli selvagem e mutantes. Em vermelho o espectro da Sfβgli selvagem e em azul do mutante indicado no gráfico. Extraído de Souza *et al.* (2016).

Foram também determinadas as constantes de Stern-Volmer, K_{sv} , para a supressão da fluorescência das Sfβgli selvagem e mutantes por acrilamida. Através desse parâmetro é possível analisar a exposição dos triptofanos da proteína, pois para suprimir a sua fluorescência a acrilamida colide com os triptofanos. Assim, quanto mais expostos na estrutura proteica estiverem os resíduos de triptofano, maior a probabilidade de colisão e supressão, e consequentemente maior o K_{sv} (equação 7). Todos os dados experimentais se ajustaram bem a

essa equação conforme visto na figura 15. Para determinar o erro experimental de K_{sv} , o ensaio com a SfbgI selvagem foi feito em triplicata e o desvio padrão calculado. O erro foi então considerado como dois desvios-padrão da média. A incerteza dos resultados para as proteínas mutantes foi considerada como igual ao erro porcentual observado nos experimentos com a SfbgI selvagem.

A figura 16 apresenta os valores de K_{sv} da SfbgI selvagem e de todos os mutantes. Como pode ser observado, considerando a incerteza experimental todos os valores são semelhantes àquele da enzima selvagem. Assim, quando analisadas por esta técnica, todas as enzimas mutantes apresentam resíduos de triptofano com exposição ao solvente semelhante à enzima selvagem, indicando que nenhuma delas esteja desnaturada ou com dobramento muito diferente.

Figura 15. Supressão da fluorescência dos resíduos triptofano de Sfβgli selvagem e mutantes utilizando acrilamida (Q). As inclinações das retas são o parâmetro K_{sv} . Extraído de Souza *et al.* (2016).

Figura 16. Valores de K_{sv} obtidos para Sfbglti selvagem e mutantes. Em cinza escuro estão os dados da Sfbglti selvagem. Em azul, enzimas com mutações em resíduos não-centrais. Em vermelho, as enzimas com mutações em resíduos centrais. Em verde, as enzimas com mutações em vizinhos de resíduos centrais.

1.4.5 Dicroísmo Circular (CD) das Sfbglti selvagem e mutantes

Em um teste adicional para avaliar se a estrutura secundária das Sfbglti mutantes são semelhantes à da selvagem, seus espectros de CD foram coletados. Como pode ser observado na figura 17, os resultados sugerem para todos os mutantes, exceto A264F, a presença de um conjunto de elementos de estrutura secundária semelhante àquele da enzima selvagem. Observa-se que o mutante A264F (vizinho do central F251) exibe um espectro de CD com

formato ligeiramente diferente da enzima selvagem, indicando que sua estrutura pode ser sutilmente diferente da enzima selvagem. No entanto, o espectro deste mutante ainda possui características de proteína enovelada. Para compreender melhor estas mudanças foi utilizando o programa de deconvolução K2D3 para analisar o espectro de CD, sendo que a proteína selvagem apresentou 33% de alfa-hélice e 25% de folhas beta, enquanto o mutante A264F apresentou uma porcentagem de alfa-hélice ligeiramente maior (37%) e a de folhas beta menor (18%). No entanto, devido a imprecisões inerentes às análises de deconvolução, torna-se difícil afirmar com base em diferenças pequenas que realmente há uma diminuição de folhas beta na estrutura da mutante A264F. Alguns trabalhos apontam, inclusive, que proteínas no estado de *molten globule* ainda mantêm seus conteúdos de estrutura secundária, indicando que mesmo formas desnaturadas podem possuir conteúdos de alfa-hélice semelhantes à forma nativa (Vassilenko e Uversky, 2002). Porém, novamente, a pequena diferença no percentual de alfa-hélices torna difícil esta afirmação. De qualquer modo, esses resultados sugerem que a mutação na posição A264F pode ter afetado o dobramento da enzima, gerando uma proteína com estrutura secundária e terciária um pouco diferente da proteína selvagem.

Vendruscolo e colaboradores (2002) mostraram que mutações em resíduos centrais da rede do estado de transição do enovelamento de proteínas afetam este processo. No entanto, no trabalho é discutida a importância da utilização da rede do estado de transição de enovelamento da proteína, e não da sua forma nativa. Dessa forma, como A264 é um resíduo vizinho a um central, o efeito sobre o enovelamento não pode ser diretamente comparado aos resultados de Vendruscolo e colaboradores (2002).

Figura 17. Espectro de CD para Sfβgli selvagem e mutantes. Os pontos vermelhos correspondem ao espectro da enzima selvagem (WT) e os azuis do mutante indicado no gráfico. As mutações marcadas por * são em resíduos não-centrais. As mutações marcadas por # estão na vizinhança do resíduo central F251. Mutações sem marcação são em resíduos centrais. Extraído de Souza *et al.* (2016).

1.4 Conclusão

Os resultados indicam que, exceto por A264F, as mutações em resíduos centrais, não-centrais e na vizinhança do central F251 não afetaram de forma significativa o enovelamento da Sfβgli. Dessa forma, é possível a caracterização dessas proteínas mutantes analisando a sua termoestabilidade (capítulo 2) e a catálise enzimática (capítulo 3).

A mutação A264F parece ter provocado uma mudança, ainda que sutil, na estrutura da Sfβgli, indicando que mudanças nesse resíduo seriam capazes de perturbar o resíduo central F251, o que pode ter resultado nessa pequena mudança estrutural da proteína. Nesta mesma linha, as mutações T245A e R320A causaram um acúmulo destes mutantes em corpos de inclusão, sendo estas respectivamente uma mutação em um resíduo central e uma em um resíduo não-central. No entanto, vale destacar que a retenção em corpos de inclusão é um fenômeno ainda não elucidado completamente, que pode ser resultado de uma conformação alternativa da proteína ou ainda resultado da maquinaria celular, não estando claramente associada a uma propriedade termodinâmica específica da proteína (Villaverde e Carrio, 2003; Ventura e Villaverde, 2006).

Desta forma, não é viável fazer uma correlação clara do resultado observado para A264F, T245A e R320 com a centralidade do resíduo mutado. Enfim, como discutido por Vedruscolo e colaboradores (2002), para melhor detalhamento da relação entre centralidade do resíduo e seu papel no dobramento da proteína, é mais adequado o uso da rede estrutural do estado de transição de dobramento da proteína, e isso somente poderá ser explorado em trabalhos posteriores.

2

TERMOESTABILIDADE E AS REDES DE ESTRUTURA PROTEICA

2.1 Introdução

2.1.1 Termoestabilidade de Proteínas

A temperatura é um fator determinante para a vida. Mesmo em regiões do planeta com temperaturas extremas, há organismos que vivem devido a sua adaptação a esses ambientes. Essa adaptação se manifesta em diversos aspectos e também pode ser observada na evolução das enzimas, que através de mutações e seleção podem ter atividade e serem estáveis em temperaturas extremas. Diversos estudos têm se dedicado a compreender como ocorre esse processo de evolução, como apontado por Arnold e colaboradores (2001) e por Somero (1995). Miyazaki e colaboradores (2000), em um estudo com a subtilisina S41 de *Bacillus* TA41 da Antártida, transformaram com somente 7 mutações esta enzima psicrófila em uma proteína com temperatura de *melting* (T_m) 23°C mais elevada. Isso demonstra que, apesar de proteínas termofílicas, mesofílicas e psicrófilas poderem estar separadas por grandes distâncias evolutivas, mudanças de termoestabilidade podem ser atingidas com poucas mutações.

Além do interesse científico nas bases moleculares da termoestabilidade, a estabilidade térmica também é bastante desejada para o uso industrial de enzimas. Enzimas termoestáveis

têm usos potenciais na produção de fármacos, como catalisadores de ensaios diagnósticos, na síntese de compostos ou na degradação de descartes, entre outros usos (Lehmann *et al.*, 2000).

Diversas características têm se mostrado importantes para a estabilização térmica de proteínas: maior número de ligações de hidrogênio e de pontes salinas, maior polaridade na superfície de proteínas e maior rigidez de regiões flexíveis (que ocorre, por exemplo, ao substituir resíduos presentes em alças por prolinas) (Lehmann *et al.*, 2000; Arnold *et al.*, 2001).

2.1.2 Termoestabilidade e Redes de Estrutura Proteica

Com o crescimento do uso das redes para descrever as proteínas, alguns grupos têm utilizado esta ferramenta para compreender a termoestabilidade de proteínas. Brinda e Vishveshwara (2005), comparando 10 proteínas termofílicas com seus análogos mesofílicos, observaram que proteínas termofílicas tendem a ter um número de resíduos altamente conectados maior do que as mesofílicas. Desta forma, as autoras propõem que a termoestabilidade das proteínas pode estar relacionada com a conectividade de sua rede de estrutura proteica.

Nessa linha, Leitner (2008) propôs uma correlação entre caminhos mínimos da rede de estrutura proteica e propagação da energia. Neste trabalho, é apontada a importância da propagação e dissipação de energia, especialmente de energia térmica, ao longo da estrutura proteica, sendo que mudanças na propagação de energia térmica poderiam resultar, inclusive, na desnaturação proteica. Dessa forma, o autor sugere que a topologia da rede pode influenciar a forma como a energia é dissipada, sendo que conexões a longo alcance, feitas por resíduos

distantes em estrutura primária, são fundamentais para que a energia térmica seja propagada muito mais eficientemente.

Utilizando estes conceitos de propagação de energia, Szalay e Csermely (2013) desenvolveram um programa chamado Turbine, que usa o chamado “princípio dos vasos comunicantes” para simular a propagação de uma perturbação ao longo de redes. Nesse trabalho, os autores observam que a modularidade da rede, *i.e.* a organização da rede em *clusters*, é algo fundamental a ser considerado ao avaliar a propagação de energia através da estrutura proteica, e que em redes com muitos *clusters* há uma maior dificuldade na propagação de energia e perturbações. De fato, como um *cluster* é altamente intra-conectado e têm poucas comunicações com outras regiões da rede, a perturbação gerada em um *cluster*, fica preponderantemente contida dentro dele, não se propagando eficientemente. Nesse sentido, nós que conectam os *clusters* uns nos outros são de fundamental importância para a propagação das perturbações, sendo que estes desempenham papel mais importante na propagação do que resíduos altamente conectados contidos dentro dos *clusters*.

Ao considerar as Redes de Estrutura Proteica, alguns trabalhos apontam que estas possuem alta modularidade, ou seja são organizadas em *clusters*, o que a princípio poderia dificultar a propagação de energia térmica (Kannan e Vishveshwara, 1999; Di Paola e Giuliani, 2015). No entanto, trabalhos têm indicado uma semelhança entre as Redes de Estrutura Proteica e um modelo de Rede chamado de **Cluster** de Percolação (Leitner, 2008; Morita e Takano, 2009). Neste modelo de rede, existem “canais” conectando um grande número de regiões da rede, fazendo com que o caminho mínimo da rede seja bastante curto e que a propagação de efeitos através dela seja facilitada. De fato, analisando as Redes de Estrutura

Proteica, temos que os resíduos centrais poderiam desempenhar papel fundamental na comunicação e fluxo de energia térmica ao longo da estrutura, pois encurtam o caminho mínimo da rede proteica, estando então a um curto número de contatos de todos os outros nós da rede. Contudo, poucos trabalhos experimentais têm buscado correlação entre a centralidade e a importância do resíduo na termoestabilidade de proteínas. Os trabalhos publicados com esse objetivo ou têm utilizado proteínas previamente caracterizadas, ou analisado proteínas *in silico*, ou ainda caracterizado proteínas sem fazer um desenho experimental que permita correlacionar efetivamente a importância dos resíduos centrais com a estabilidade térmica de proteínas.

Nesse sentido, este capítulo desta tese tem como objetivo avaliar a correlação entre a centralidade dos resíduos da rede de estrutura proteica da beta-glicosidase de *Spodoptera frugiperda* (Sf β gli) com a sua importância na estabilidade térmica desta proteína. A propriedade de termoestabilidade foi escolhida, pois serve como um indicador da influência do resíduo central na estrutura proteica como um todo. Para isso, utilizamos as proteínas Sf β gli com mutações construídas e produzidas segundo o capítulo 1, sendo que dentre essas proteínas existem dois grupos: 1 - proteínas mutadas de forma a perturbar os resíduos centrais da Sf β gli, quer seja por mutações nos próprios resíduos centrais da rede, quer seja por mutações que perturbam o ambiente no qual estão inseridos os resíduos centrais, e 2 - mutações controle que, a princípio, não perturbam resíduos centrais.

2.2 Materiais e Métodos

Para coleta de dados, foram utilizadas amostras produzidas e purificadas segundo os passos descritos no capítulo 1. Foram utilizadas amostras produzidas em bactérias *BL21*

gold(DE3). Além disso, após a purificação, foi feita uma cromatografia de dessalinização, trocando o tampão de eluição para fosfato de potássio 20 mmol·L⁻¹ pH 6,0.

2.2.1 Determinação da Temperatura de *melting* (T_m) por Dicroísmo Circular (CD)

A T_m das Sfβgli selvagem e mutantes foi determinada através da análise de dados de CD. Para isso a elipsidade foi medida em função do aumento da temperatura, utilizado o espectropolarímetro J-720 (JASCO, Inc), coletando a elipsidade a 222 nm, 215 nm e 208 nm. Foi utilizada uma rampa de temperatura de 20 a 90°C, com aumento de temperatura de 0,5 °C/min. A elipsidade foi coletada após 5 segundos da temperatura alvo ser atingida. Os valores de elipsidade foram transformados em elipsidade molar utilizando a Equação 8 (capítulo 1). Após isso, gráficos de elipsidade molar *versus* temperatura foram plotados. Para determinar a T_m , as curvas obtidas foram ajustadas para a equação sigmoidal (9), utilizando o programa Origin 8.5.1.

$$y = (A2x + B2) + \frac{(A1x + B1) - (A2x + B2)}{1 + e^{\frac{x - T_m}{dx}}} \quad \text{Equação 9}$$

Onde $A1x + B1$ é a reta que descreve a região antes da transição; $A2x + B2$ é a reta que descreve a região pós-transição e dx é o parâmetro que descreve o formato da curva na região de transição.

2.2.2 Differential Scanning Fluorimetry (DSF)

Para avaliar a estabilidade térmica foi utilizado o método de *Differential Scanning Fluorimetry* (DSF), também chamado de *ThermoFluor* ou *Thermal Shift Assay* (TSA) (Pantoliano *et al.*, 2001).

Nos ensaios foram utilizadas concentrações em uma faixa de 25 ng a 1,0 µg de Sfβgli em um volume final de 22,5 µL de tampão fosfato de potássio 20 mmol·L⁻¹ pH 6,0 em tubos de PCR ópticos. Então, foram adicionados 2,5 µL de corante SyproOrange® (Sigma-Aldrich) previamente diluído 500×. Os tubos foram, então, fechados com tampas para tubos de PCR ópticos. Foram feitas rampas de aumento de temperatura com incremento de 0,5 % por etapa em um equipamento de PCR em tempo real (*AppliedBioscience* ABI7500) no modo de *melting temperature*, sendo que as amostras foram excitadas com uma lâmpada de tungstênio-halogênio e as leituras de fluorescência foram coletadas com o filtro 2 e 3 (550 nm e 580 nm, respectivamente).

Após os ensaios, foram plotados o sinal de fluorescência *versus* temperatura para todas as Sfβgli selvagem e mutantes. A temperatura de *melting* foi calculada através do pico da primeira derivada da curva obtida. Quando necessário, a curva foi atenuada para suavizar o ruído. Estes procedimentos foram realizados utilizando o programa Origin 8.5.1. Todas os experimentos foram realizados em triplicata, e o erro da T_m foi estimado através do desvio padrão dessa triplicata.

2.2.3 Cinética de desnaturação utilizando o corante *SyproOrange*[®]

Para avaliar a estabilidade térmica das proteínas selvagem e mutantes foi executado ensaio de cinética de desnaturação térmica utilizando a sonda *SyproOrange*[®]. Para isso, foram preparados 1,0 mL de solução contendo 40 µg de proteína em tampão fosfato de potássio 10 mM pH 6,0, contendo *SyproOrange*[®] 4000x diluído (diluição final). Para minimizar a precipitação do corante, a diluição foi feita no momento da medição. Essa amostra foi incubada a 45 °C (ou outras temperaturas desnaturantes) em um fluorímetro (*Hitachi f4500*) acoplado a um banho termostatizado e com circulação de água (*Pharmacia Biotech*). Simultaneamente ao processo de incubação, foram coletadas durante 1 hora as emissões fluorescências a 570 nm com excitação de 472 nm utilizando o modo de leitura contínua do equipamento (*Time Scan Mode*). A curva resultante foi ajustada a equação de cinética de primeira ordem (equação 10) utilizando o software Origin 8.5.1. Além disso, os dados foram transformados em valores relativos foram linearizados plotando-se o $\ln(D\%)$ versus tempo, onde D% é a porcentagem de enzima na forma desnaturada, calculada segundo a equação 11.

$$I(t) = (I(\text{max}) - I(0)) \cdot (1 - e^{-k_{\text{desn}}t}) + I(0) \quad \text{Equação 10}$$

$$D\% = \frac{I(t) - I(0)}{I(\text{max}) - I(0)} \quad \text{Equação 11}$$

Onde $I(t)$ é a intensidade de fluorescência no tempo t , $I(0)$ é a intensidade de fluorescência inicial da amostra, e $I(\text{max})$ é a intensidade de fluorescência quando t tende a altos valores e k_{desn} , a constante de desnaturação térmica.

2.2.4 Plote de Eyring e cálculo da energia de ativação da desnaturação

Para avaliar a energia de ativação da desnaturação da Sfβgli selvagem e mutantes foram construídos plotes de Eyring. Para isso, foram feitos ensaios de desnaturação utilizando SyproOrange (item 2.2.3) em diferentes temperaturas conforme descrito no item 2.2.3. Ao final, gráficos de $\ln(k_{\text{desn}}/T)$ versus $1/T$ (onde T é a temperatura em Kelvin e k_{desn} é a constante cinética de desnaturação) foram plotados. Conforme visto na equação 12, a partir da inclinação da reta observada no plote de Eyring foi calculado o $\Delta H^{\ddagger}_{\text{desn}}$ da reação de hidrólise do substrato, enquanto que o intercepto permitiu o cálculo da $\Delta S^{\ddagger}_{\text{desn}}$. Com estes parâmetros, foi calculado o $\Delta G^{\ddagger}_{\text{desn}}$ a 45 °C.

$$\ln\left(\frac{k_{\text{desn}}}{T}\right) = -\frac{\Delta H^{\ddagger}}{R} \cdot \frac{1}{T} + \ln \frac{k_B}{h} + \frac{\Delta S^{\ddagger}}{R} \quad \text{Equação 12}$$

Onde T é a temperatura em Kelvin, R é a constante dos gases ideais, k_B é a constante de Boltzmann e h é a constante de Planck.

2.2.5 Cinética de inativação acompanhada pela atividade enzimática remanescente

Para avaliar a estabilidade térmica das proteínas selvagem e mutantes, foi executado ensaio de cinética de inativação térmica a 45 °C. Para isso, a proteína foi diluída para uma concentração adequada ao ensaio de atividade enzimática, utilizando tampão citrato-fosfato 100 mmol·L⁻¹ pH 6,0, para um volume final de 3,0 mL. Essa amostra foi, então, incubada a 45 °C em um banho termostatizado (*Pharmacia Biotech*). De tempos em tempos, alíquotas de aproximadamente 200 μL dessa amostra foram removidas do tubo, e colocadas em gelo. Com

essas alíquotas, foram feitos ensaios de atividade enzimática, adicionando-se 50 μL de amostra a 50 μL de substrato *p*-nitrofenil- β -D-fucopiranosídeo (pNPFuc) 2 $\text{mmol}\cdot\text{L}^{-1}$ e incubando-se a mistura a 30 °C. Depois de 5, 10, 15 ou 20 minutos, a reação foi interrompida adicionando-se 100 μL de carbonato de sódio 0,5 $\text{mol}\cdot\text{L}^{-1}$, e as absorbâncias a 415 nm (Abs_{415}) dos meios reacionais foram lidas em placas de 96 poços. Gráficos de Abs_{415} versus tempo foram plotados, onde a inclinação da reta obtida é proporcional à atividade enzimática. Ao final, a atividade porcentual com relação à amostra não incubada a 45 °C foi calculada para cada uma das alíquotas. Gráficos do logaritmo da atividade porcentual versus o tempo de incubação a 45 °C foram plotados, sendo que a constante de inativação térmica é a inclinação dessa curva.

2.3 Resultados e Discussão

Para avaliar o efeito das mutações em resíduos centrais e sua vizinhança na termoestabilidade foram determinadas as temperaturas de *melting* (T_m), as constantes de velocidade de desnaturação e inativação a 45 °C (k_{desn} e k_{inat}) e a energia de ativação do processo de desnaturação ($\Delta G^{\ddagger}_{\text{desn}}$) para todas as Sf β gli mutantes, sendo estes valores comparados com aqueles da Sf β gli selvagem.

2.3.1 Determinação da T_m das Sf β gli selvagem e mutantes

As elipsidades molares a 222 nm, 215 nm e 208 nm foram coletadas para todas as Sf β gli (selvagem e mutantes) ao longo de um gradiente de temperatura de 20 a 90 °C. Através disso foi possível monitorar variações de estrutura secundária das proteínas com o aquecimento, e determinar a temperatura de desnaturação (T_m) destas enzimas. Na figura 18, observa-se que as curvas de *melting* utilizando a elipsidade molar a 222 nm e 215 nm apresentam um formato

sigmoidal. Enquanto isso, a curva a 208 nm não apresenta este formato, possivelmente porque a desnaturação não provoca uma grande mudança nessa região do espectro, atuando mais na estrutura secundária que gera sinal nos outros comprimentos de onda (222 e 215 nm). Comparando a T_m calculada utilizando os comprimentos de onda de 222 nm e 215 nm, observa-se que o valor é aproximadamente igual. Isso é um indício de que a mudança observada nestes dois comprimentos de onda reflete a mesma alteração estrutural. Dessa forma, para análise dos resultados das proteínas mutantes, foram considerados os perfis de desnaturação a 222 nm, que apresentaram uma mudança de elipsidade molar de maior intensidade. Os resultados do perfil de desnaturação das SfbgIi mutantes acompanhados pela elipsidade a 222 nm estão na figura 19.

Figura 18. Perfil de desnaturação térmica da SfbgIi selvagem acompanhado por CD. Em A, elipsidade molar a 222 nm; em B, a 215 nm; e em C, a 208 nm.

Figura 19. Desnaturalização térmica dos mutantes acompanhada através de CD. Leituras de elipsidade molar foram coletadas em 222 nm. Os pontos vermelhos são os dados coletados e a linha preta o ajuste na equação descrita em materiais e métodos. Os dados para o mutante A264F foi melhor ajustado em uma equação de reta.

Além disso, foram feitos ensaios de *Differential Scanning Fluorimetry* (DSF) de todos os mutantes e da enzima selvagem. Neste experimento, a amostra é incubada com um corante, SyproOrange. Este corante exibe baixa fluorescência em meio aquoso, e alta fluorescência em meio mais apolar. Desta forma, ao colocar em um tubo a proteína de interesse e o corante, o

sistema emitirá baixa fluorescência (no comprimento de onda do corante). No entanto, atingindo uma temperatura tal que a proteína passa a desnaturar, os grupos hidrofóbicos do interior da proteína tornam-se expostos, e o corante passa a interagir com eles, aumentando sua fluorescência. Dessa forma, o aumento da fluorescência torna-se um indicador da desnaturação da proteína, e a T_m é a inflexão da curva observada. Por fim, em temperaturas muito elevadas, a proteína pode sofrer processo de agregação, o que diminui a interação do corante com os grupos hidrofóbicos da proteína, promovendo uma redução da fluorescência no final da curva (Niesen *et al.*, 2007).

Na figura 20 está o resultado observado para cada uma das Sfbg1i mutantes em comparação com o resultado observado para a proteína selvagem. Observa-se que alguns mutantes apresentam perfis de DSF diferentes da proteína selvagem, ou por apresentar transição em uma temperatura diferente, ou por apresentar a linha de pré-transição mais elevada do que a selvagem.

Segundo Luan e colaboradores (2014), a intensidade de fluorescência na fase pré-transição também é bastante importante, estando relacionada com a exposição de grupos hidrofóbicos e, desta forma, com a estabilidade da proteína em baixas temperaturas.

Para avaliar se os resultados obtidos eram reprodutíveis, os experimentos com a Sfbg1i selvagem e mutantes foram repetidos 3 vezes. Foi observado que a T_m é bastante reprodutível, mas a intensidade de fluorescência detectada no experimento variou significativamente, provavelmente porque ela depende da concentração de corante e da concentração de proteína. Além disso, o ensaio é feito em volumes pequenos, o que potencializa os erros relacionados à concentração. Adicionalmente fatores, como tampão, pH e força iônica podem provocar

mudanças no perfil de desnaturação de proteínas. Para minimizar os efeitos destes fatores, vale destacar que os experimentos de DSF foram feitos sempre com o mesmo tampão e que as proteínas foram diluídas de forma a ficarem com a mesma concentração final. Com base nestes comentários, optou-se por normalizar os dados e avaliar a T_m , conforme a figura 20, deixando de lado a intensidade na pré-transição.

Mesmo com os cuidados comentados acima, para nos certificarmos que as diferenças de T_m observadas não estavam associadas a mudanças na concentração da Sfβgli, foram feitos experimentos de DSF utilizando diferentes concentrações de proteína. O resultado obtido (Figura 21) revela que a T_m é independente da concentração de Sfβgli, o que confirma que as diferenças de T_m são devidas às mutações. Além disso, como a desnaturação térmica da Sfβgli não depende da sua concentração, a etapa limitante do processo de desnaturação não é a agregação das proteínas, que é fortemente dependente da concentração.

Para relacionar o efeito observado com a posição das mutações, foram feitos gráficos comparativos entre as T_m de todas as Sfβgli (Figura 22). Observa-se que os valores de T_m obtidos por CD quanto por DSF (Tabela 3; Figura 22) concordam entre si. Esse resultado sugere que a desnaturação térmica detectada por ambas as técnicas deve se tratar do mesmo processo, mesmo que elas detectem propriedades diferentes das proteínas, sendo que o CD detecta principalmente mudanças na estrutura secundária de proteínas, e o DSF detecta mudanças de estrutura terciária.

Figura 20. Comparação entre dos perfis obtidos por DSF das Sfβgli mutantes com a selvagem. Em vermelho, a curva da proteína selvagem e em azul, do mutante indicado no gráfico. Extraído de Souza *et al.* (2016).

Figura 21. Curvas de DSF de todos os mutantes de SfbgI. As curvas em azul foram coletadas usando 1 μg de SfbgI; as curvas em vermelho com 500 ng, em verde usando 250 ng e em roxo, 125 ng. Adaptado de Souza *et al.* (2016).

Figura 22. T_m das SfbgI selvagem e mutantes. Em A, mutações que não perturbam os resíduos centrais (azul) e mutações dirigidas diretamente aos resíduos centrais (vermelho). Em B, mutações na vizinhança do resíduo central F251. O retângulo representa resíduos posicionados em um alfa-hélice, enquanto a seta os resíduos em fita beta. Não foram detectadas transições e portanto a T_m para A264F não foi calculada. Barras hachuradas representam dados obtidos por DSF, enquanto barras vazias representam dados obtidos por CD. Figura extraída de Souza *et al.* (2016).

Tabela 3. Valores de T_m da SfbgI selvagem e mutantes determinados através de dicroísmo circular (CD) e DSF. Extraído de Souza *et al.* (2016).

Enzima	T_m (°C) ^a	T_m (°C) ^b
Selvagem	46.0 ± 0.2	45.0 ± 0.7
L100A	46.4 ± 0.9	47.6 ± 0.9
L350A	43.5 ± 0.6	43.8 ± 0.9
L389A	46.4 ± 0.3	46.6 ± 0.7
Y390A	43 ± 2	45.1 ± 0.6
N391A	44.4 ± 0.3	46.0 ± 0.9
S247A	45.7 ± 0.4	45.3 ± 0.4
N249A	38.5 ± 0.5	39 ± 1 [*]
F251A	42.2 ± 0.6	42.4 ± 0.3 [*]
F334A	42.9 ± 0.5	42.5 ± 0.7 [*]
S358A	42.7 ± 0.8	42 ± 1 [*]
K366A	41.5 ± 0.7	41 ± 1 [*]
Y420A	37.0 ± 0.6	37.2 ± 0.3 [*]
<i>D260A</i>	39.6 ± 0.3	40.8 ± 0.6 [*]
<i>E261A</i>	45.2 ± 0.3	46.2 ± 0.9
<i>M262F</i>	45.7 ± 0.2	45 ± 1
<i>A263F</i>	39.6 ± 0.4	40.6 ± 0.4 [*]
<i>A264F</i>	ND	ND
<i>E265A</i>	44.9 ± 0.4	44.9 ± 0.3
<i>L335A</i>	38.5 ± 0.2	40 ± 1 [*]
<i>V336F</i>	30 ± 2	34 ± 2 [*]
<i>S337F</i>	39.1 ± 0.4	40.42 ± 0.01 [*]

^a T_m foi monitorada através de CD. Incertezas foram determinadas através do ajuste a curva sigmoidal pelo Origin.

^b T_m foi determinada por DSF. Incertezas foram determinadas através do desvio padrão de 3 experimentos independentes.

^{*} T_m significativamente diferente da enzima selvagem ($P \leq 0.01$).

Mutações em resíduos centrais são apresentadas em negrito. Mutações na vizinhança do resíduo central F251 estão em itálico. Mutações em resíduos não-centrais estão em fonte normal.

ND, como a desnaturação do mutante A264F envolve múltiplos intermediários, não foi determinada uma única T_m .

Na figura 22A observa-se, de uma maneira geral, que as mutações que substituíram os resíduos centrais por alanina (barras em vermelho) resultaram em proteínas com T_m menor do que SfbgI selvagem. Por outro lado, as mutações que não perturbam os resíduos centrais (barras em azul) não causaram efeitos na T_m das proteínas. Isso sugere que a troca dos resíduos centrais por alanina, o que resulta em ruptura de grande parte das interações desses resíduos com sua vizinhança, promove um efeito na estrutura proteica como um todo reduzindo sua termoestabilidade. De forma consistente com esse resultado, a figura 22B também indica uma redução de termoestabilidade nas SfbgI mutantes que contém perturbações próximas ao resíduo central F251.

Curiosamente, a redução de termoestabilidade parece estar associada ao posicionamento da cadeia lateral dos resíduos mutados com relação ao F251. Analisando inicialmente a fita beta contendo os resíduos L335, V336 e S337 (Figura 23A), observa-se que a mutação que provocou maior efeito na T_m foi V336F (Figura 22B), sendo que esta valina apresenta cadeia lateral apontada para o resíduo central F251, enquanto os demais integrantes dessa fita beta (L335 e S337), apesar de próximos, não têm cadeia lateral na face da fita voltada para F251. Similarmente, ao analisar a alfa hélice contendo os resíduos D260, E261, M262, A263, A264 e E265 (Figura 23B), observa-se que as mutações em resíduos com cadeia lateral apontada na direção do central F251 (D260A, A263F e A264F) foram as que resultaram em SfbgI mutantes com menor T_m , enquanto que as mutações em resíduos com cadeia lateral apontada em direção oposta ao F251 (M262A e E265A) não afetam a T_m da proteína. Esses resultados sugerem que os efeitos observados ocorrem via o resíduo F251, indicando que a diminuição da estabilidade térmica se deve, de fato, à perturbação de resíduos centrais.

Figura 23. Vizinhança do resíduo F251. Em A, resíduos mutados para perturbar o F251. Em B, vista da alfa-hélice evidenciando os resíduos cuja cadeia lateral está apontada para o F251. Figura extraída de Souza et al., (2016).

Além da conclusão geral apresentada acima, um dado específico e interessante que pode ser observado nas figuras 19, 20 e 21 é que não há T_m para Sfβgli A264F. De fato, ao analisar as figuras (repetidas nos painéis D e E da Figura 24; abaixo), observa-se que tanto pela técnica de DSF quanto por CD, as medidas coletadas, que refletem os perfis de desnaturação, têm um comportamento linear em função da temperatura, e não sigmoidal. Não há, portanto, uma transição clara entre forma nativa e forma desnaturada. Ao observar esse perfil, pode-se pensar inicialmente que a Sfβgli A264F já esteja na forma desnaturada desde o início do experimento. No entanto, na figura 24F, observa-se que a Sfβgli A264F possui um espectro de dicroísmo circular característico de uma proteína enovelada, apresentando apenas algumas pequenas diferenças em relação à Sfβgli selvagem (Figura 24C). Além disso, uma comparação entre os espectros de CD coletados a 20 °C e a 90 °C da proteína selvagem e deste mutante mostram claramente que, como para a enzima selvagem, o espectro da Sfβgli A264F muda de 20 °C para 90 °C, sugerindo que há um processo de desnaturação associado ao aumento de temperatura.

Além disso, os espectros de CD coletados a 90 °C para a SfbgI selvagem e A264F são iguais, sugerindo que ambas desnaturam e atingem um estado semelhante. Estes resultados indicam que o mutante A264F não está totalmente desnaturado no início do experimento.

Figura 24. Desnaturação térmica da SfbgI selvagem e do mutante A264F. Em A, desnaturação térmica acompanhada por CD da SfbgI selvagem. Em B, desnaturação por DSF da SfbgI selvagem. Em C, espectro de CD da SfbgI selvagem a 20°C (pontos azuis) e a 90°C (pontos verdes). Em D, desnaturação acompanhada por CD da SfbgI A264F. Em E, desnaturação acompanhada por DSF da SfbgI A264F. Em F, espectro de CD da SfbgI A264F a 20°C (pontos azuis), e a 90°C (pontos verdes). Extraído de Souza *et al.* (2016).

Através da figura 25, nota-se que a mutação A264F afeta fortemente o resíduo central F251, pois o resíduo de alanina nessa posição está próximo ao resíduo F251 (F251 C_β – A264 C_β: 5.3 Å), com a cadeia lateral apontada para este resíduo central. Dessa forma, a substituição A264F deve ter provocado uma grande perturbação do resíduo F251, o que resulta neste grande efeito sobre o processo de desnaturação, que deixou de ser cooperativo.

Figura 25. Efeito da mutação A264F sobre o resíduo central F251. Em A, a região do resíduo F251 na enzima selvagem. Em B, a região do resíduo F251 com a mutação A264F. Em vermelho está o resíduo F251, e em azul, o resíduo na posição 264.

Na literatura, há um estudo utilizando pinças-ópticas de molécula única, que investigou a natureza cooperativa de desnaturação de proteínas, sendo que os resultados indicaram que falta de acoplamento entre domínios ou regiões estruturadas pode remover a cooperatividade dos processos de desnaturação e enovelamento proteico (Shank *et al.*, 2010). Considerando a abordagem aqui empregada de representação da estrutura de proteínas como redes, o alto coeficiente de *clustering* presente nas redes proteicas indica que elas são organizadas em regiões, grupos de aminoácidos, com alta conectividade local, as quais são conectadas entre si através dos centrais (Brinda e Vishveshwara, 2005). Dessa forma, mutações como A264F, que perturba drasticamente um resíduo central, poderiam remover ou reduzir o acoplamento entre os *clusters* de resíduos da estrutura proteica, abolindo a cooperatividade na sua desnaturação.

2.3.2 Cinética de desnaturação térmica acompanhada por Sypro Orange

Para compreender melhor a estabilidade térmica das Sfβgli selvagem e mutantes, foram realizados experimentos de cinética de desnaturação térmica a uma temperatura fixa. Nesses experimentos a cinética do processo de desnaturação é o fator avaliado, diferentemente dos

anteriores feitos com rampa de temperatura onde a termodinâmica do processo foi analisada. Dessa forma, estes experimentos de cinética podem trazer informações complementares às informações anteriores. Para executar o experimento, inicialmente, a metodologia descrita pelo item 2.2.3 foi padronizada. Nesta metodologia, a proteína é incubada com o corante Sypro Orange em uma cubeta de fluorímetro. Como o corante está em meio aquoso, ele exibe uma baixa fluorescência. A partir de um banho acoplado ao fluorímetro, uma temperatura é ajustada, sendo ela alta suficiente para desnaturar a proteína. Conforme a proteína desnatura, ela expõe grupos apolares, permitindo interação do corante com esses grupos hidrofóbicos, o que resulta em aumento da fluorescência ao longo do tempo. Dessa forma, o corante age como uma sonda para avaliar a cinética do processo de desnaturação.

A partir disso, inicialmente, todas as SfbgIi selvagem e mutantes foram desnaturadas a 45 °C e a cinética do processo foi acompanhada (Figura 26). O resultado foi linearizado através de um plote semilogarítmico, e desta forma as constantes cinéticas de desnaturação (k_{desn}) foram determinadas. A figura 27 compara os valores de k_{desn} das SfbgIi selvagem e mutantes. Observa-se que as mutações em resíduos centrais (barras vermelhas) promovem um aumento no k_{desn} , enquanto que os resíduos em não-centrais (barras azuis) praticamente não afetam esta constante de velocidade (Figura 27A), indicando que as mutações em resíduos centrais tornaram o processo de desnaturação a 45 °C mais rápido quando comparadas as mutações em resíduos não-centrais. Por outro lado, analisando a figura 27B, observa-se que apenas algumas das mutações que perturbam o resíduo central F251 por mudança em sua vizinhança causaram aumento de k_{desn} . Ao observar a posição relativa de F251 e dos resíduos vizinhos (Figura 23), nota-se que as mutações que aumentaram os valores de k_{desn} afetaram resíduos próximos ao

F251 e com cadeia lateral apontada para este resíduo central. Observa-se que a mutação V336F, em uma valina que tem cadeia lateral apontada para o resíduo F251 ($F251 C_{\beta} - V336 C_{\gamma}$: 3.9 Å), teve maior efeito do que as mutações L335A e S337F, que atingem resíduos com cadeia lateral direcionada em sentido praticamente oposto. Já ao observar a alfa-hélice vizinha à F251, temos que as mutações D260 e A263F causaram maiores efeitos sobre k_{desn} , sendo que se considerarmos que a alfa-hélice dá uma volta em 4 resíduos, temos que D260 e A263F estão com a cadeia lateral na mesma face da alfa-hélice que está voltada para o resíduo central F251 (Figura 23). Nesta mesma face da alfa-hélice encontra-se também a mutação A264F que resultou em uma proteína cuja fluorescência praticamente não mudou a ao longo do processo de desnaturação, já apresentando um valor razoavelmente alto em $t=0$, o que impediu o cálculo da constante de velocidade de desnaturação por essa técnica. A mutação A264F é justamente a mais próxima do resíduo F251, indicando que ela pode ter promovido grande perturbação do resíduo central, mudando o mecanismo de desnaturação de Sfβgli. Esta observação é coerente com os experimentos de medida de T_m do mutante A264F discutidos no item anterior.

Figura 26. Perfil cinético da desnaturação térmica a 45°C da Sfβgli selvagem e mutantes. As linhas azuis correspondem aos dados experimentais coletados de forma contínua. As linhas pretas correspondem ao ajuste linear correspondente.

Figura 27. Constantes cinéticas de desnaturação térmica (k_{desn}) a 45 °C das SfbgI selvagem e mutantes. Em A, gráfico comparando o efeito entre as mutações em resíduos centrais (barras vermelhas) e em resíduos não-centrais (barras azuis). Em B, resultados das mutações na vizinhança do resíduo central F251. Experimentos de desnaturação térmica foram feitos a 45 °C. Extraído de Souza et al. (2016).

Dessa forma, observa-se que em linhas gerais o efeito das mutações na cinética de desnaturação a 45°C é bastante semelhante ao efeito termodinâmico observado pela mudança de T_m .

Para aprofundar a comparação dos resultados entre os mutantes (que apresentaram diferentes valores de T_m , sendo alguns bem abaixo de 45 °C), o experimento de cinética de desnaturação das SfbgI selvagem e mutantes foi realizado em diferentes temperaturas e foram

construídos plots de Eyring (equação 12, Figura 28), relacionando a temperatura de desnaturação com a sua respectiva constante cinética do processo (k_{desn} ; Figura 27). Esse experimento não foi realizado com os mutantes A264F e N391A devido a baixas concentrações de proteína associadas a baixa mudança na intensidade de fluorescência durante o experimento.

Figura 28. Plotes de Eyring da desnaturação térmica da SfbgI selvagem e mutantes. Experimentos foram realizados na faixa de 30 °C a 50 °C. Extraído de Souza *et al.* (2016).

Conforme o esperado, o plote de Eyring de todas as SfbgI mutantes analisadas apresentou comportamento linear (Figura 28), e com isso o valor de $\Delta G_{\text{desn}}^{\ddagger}$, bem como os valores de $\Delta H_{\text{desn}}^{\ddagger}$ e $\Delta S_{\text{desn}}^{\ddagger}$ a 45 °C foram calculados. Na tabela 4 estão os valores encontrados, bem como a diferença de valores observada entre os mutantes e a proteína selvagem. Nota-se que os valores de entalpia e entropia da barreira do processo de desnaturação sofrem mudanças de energia muito semelhantes uma com as outras para um mesmo mutante, sendo então os valores de $\Delta \Delta H_{\text{desn}}^{\ddagger}$ e $\Delta \Delta S_{\text{desn}}^{\ddagger}$ praticamente iguais (Tabela 4). Esse fenômeno, chamado de efeito

de compensação entropia-entalpia, é frequentemente relatado em sistemas dominados por forças fracas (Dunitz, 1995). Adicionalmente, de acordo com o esperado, os valores de $\Delta G_{\text{desn}}^{\ddagger}$ se comportam da mesma forma do que o valor de k_{desn} , tendendo a serem menores para as proteínas com mutações em resíduos centrais, reforçando a hipótese de que perturbações diretamente dirigidas aos resíduos centrais têm efeitos na estabilidade térmica das proteínas.

Tabela 4. Parâmetros termodinâmicos da desnaturação térmica da SfbgIi selvagem e mutantes a 45°C. As mutações indicadas em negrito são em resíduos centrais e as indicadas em itálico são na vizinhança do resíduo F251. Mutações em fonte normal são em resíduos não-centrais. Wild-type, selvagem. Extraído de Souza *et al.* (2016)

Enzyme	$\Delta G^{\ddagger}$ (kJ·mol ⁻¹)	$\Delta H^{\ddagger}$ (kJ·mol ⁻¹)	$\Delta\Delta S^{\ddagger}$ (kJ·mol ⁻¹)	$\Delta\Delta H^{\ddagger}$ (kJ·mol ⁻¹)	$T\Delta\Delta S^{\ddagger}$ (kJ·mol ⁻¹)
Wild-type	96.7	355	258	–	–
L100A	100.1	221	121	134	–137
L350A	93.8	256	162	–99	–96
L389A	99.5	338	239	–17	–20
Y390A	95.3	302	207	–53	–52
S247A	96.5	281	184	–74	–74
N249A	92.8	201	108	–154	–150
F251A	93.9	218	124	–137	–134
F334A	92.1	302	210	–53	–48
S358A	94.1	215	121	–140	–137
K366A	91.5	311	219	–44	–39
Y420A	90.8	199	108	–156	–150
<i>D260A</i>	93.9	183	89	–172	–169
<i>E261A</i>	98.9	242	143	–113	–115
<i>M262A</i>	98.9	382	283	27	25
<i>A263F</i>	94.9	184	89	–171	–169
<i>A264F</i>	ND	ND	ND	ND	ND
<i>E265A</i>	98.5	213	114	–142	–144
<i>L335A</i>	93.2	230	137	–125	–121
<i>V336F</i>	87.8	215	127	–140	–131
<i>S337F</i>	93.6	224	130	–131	–128

Finalmente, como as mutações que perturbam os resíduos centrais resultaram em alterações no $\Delta G_{\text{desn}}^\ddagger$ e também em diminuição da T_m , (ver item 2.3.1), é possível que tenha ocorrido uma diminuição na estabilidade da forma nativa destas proteínas mutantes, pois esta forma é comum aos cálculos de $\Delta G_{\text{desn}}^\ddagger$ e ΔG_{desn}^0 conforme ilustrado no esquema da figura 29.

Figura 29. Perfil de energia dos mutantes durante o processo de desnaturação. Em preto, o perfil energético das enzimas mutantes em relação ao perfil da selvagem (vermelho). Observa-se, portanto, perda de estabilidade e diminuição da barreira.

2.3.3 Cinética de Inativação térmica da Sfβgli selvagem e mutantes

Analisando os resultados de inativação térmica da atividade enzimática da Sfβgli selvagem e mutantes de resíduos centrais (S247A, N249A, F251A, F334A, S358A, K366A e Y420A), Ikegami (2013) observou que, muitas vezes, as mutações dirigidas diretamente aos resíduos centrais modificavam a cinética de inativação, fazendo com que ela não fosse completamente linearizada por gráficos semilogarítmicos, sendo que a atividade das enzimas mutantes nesse tipo de gráfico decaía como uma curva resultante da combinação de duas retas com inclinações diferentes (Figura 30). Normalmente a inativação da enzima selvagem segue uma cinética de primeira ordem e, portanto, poderia ser linearizada através do gráfico semilogarítmico.

Figura 30. Inativação térmica a 45 °C das Sfβgli selvagem e F251A. Extraído da tese de C. M. Ikegami (2013).

Para aprofundar o entendimento deste fenômeno e avaliar se também ocorria com mutações em resíduos não-centrais e com mutações nos arredores de um resíduo central, o

experimento de cinética de inativação térmica foram realizados com as Sfβgli mutantes D260A, E261A, M262A, A263F, A264F, E265A, L335A, S337F, L389A e Y390A. As Sfβgli V336F, N391A e L100A não foram caracterizadas por essa técnica devido a dificuldades ocasionadas pela baixa concentração associadas à baixa atividade.

Além disso, conforme observado na figura 31, os resultados para as enzimas mutantes são parecidos com os de Ikegami (2013), sendo que nos perfis de inativação a atividade decai inicialmente como uma reta com alta inclinação e depois assume uma segunda reta com inclinação menor. As enzimas contendo as mutações A264F, L335A e S337F foram as únicas a apresentar comportamento diferente, sendo que inativaram de acordo com um mecanismo de primeira ordem ao longo de toda faixa de tempo testada. Inclusive no caso de A264F e L335A a inativação tornou-se mais rápida do que a própria enzima selvagem.

Deste modo, buscando analisar sistematicamente este padrão de inativação curvilíneo, optou-se por recorrer a modelos cinéticos nos quais os dados de inativação das Sfβgli mutantes e selvagem poderiam se ajustar.

Figura 31. Inativação térmica das *Sfβgli* selvagem e mutantes a 45°C. Em azul, os dados correspondentes a *Sfβgli* mutante e em vermelho, os dados da *Sfβgli* selvagem.

Um deles supõe que nas condições de inativação, com a temperatura constante de 45 °C, a desnaturação de Sfβgli pode estar ocorrendo de forma reversível (Equação 13). Dessa forma, a Sfβgli se desnatura até que o ΔG do sistema torne-se zero, sendo então atingido um equilíbrio, onde as velocidades de desnaturação e renaturação tornam-se iguais e, portanto, a concentração de Sfβgli ativa deixa de diminuir. Esse fenômeno já foi observado para outras proteínas, e é possível tratar os dados de acordo com uma cinética de primeira ordem reversível (equação 14) adaptada de (Levine, 2009).

$$N \rightleftharpoons D \quad \text{Equação 13}$$

$$A = (1 - A_{eq})e^{-kt} + A_{eq} \quad \text{Equação 14}$$

Onde A é a atividade residual relativa no tempo t (atividade após o tempo t de inativação dividida pela atividade em t = 0), A_{eq} é a atividade residual relativa no equilíbrio, e k é a constante de velocidade de inativação, que é a diferença das velocidades direta e reversa do equilíbrio.

Nesse modelo temos que o patamar alcançado após a fase inicial de desnaturação visto na Figura 31 se referiria à concentração de Sfβgli ativa presente no equilíbrio. Com isso, mudanças no patamar decorrente das mutações revelariam que há mudanças no equilíbrio entre os estados nativo e desnaturado de Sfβgli. No entanto, observando a figura 31, verifica-se que os gráficos com algumas Sfβgli mutantes, como por exemplo o mutante A263F, não possuem um patamar de atividade relativa paralelo ao eixo x, como esperado em um equilíbrio, mas sim uma segunda fase de inativação com menor velocidade. Logo, este modelo cinético não parece ser adequado para descrever os dados experimentais aqui obtidos.

Um outro modelo parte da hipótese de que a desnaturação da Sfβgli pode não seguir o mecanismo de uma única etapa. Nesse caso, haveria a formação de um intermediário, o qual teria atividade menor do que da Sfβgli nativa. Nesse caso, a fase inicial de inativação, com maior velocidade, poderia estar relacionada com a forma nativa da Sfβgli se transformando no intermediário parcialmente ativo e a segunda etapa, mais lenta, poderia estar relacionada com a forma intermediária da Sfβgli se transformando na forma desnaturada final. Um resultado que corrobora esta hipótese de um intermediário ativo é que a Sfβgli A264F, apesar de ter espectro de CD diferente da Sfβgli selvagem, mantém atividade enzimática, o que indica que mesmo uma forma diferente da nativa pode ter atividade enzimática residual. Com isso, mudanças na constante de velocidade da primeira fase estariam associadas à formação desse intermediário parcialmente ativo, e alterações na segunda fase de inativação indicariam mudanças no perfil de desnaturação do intermediário para a forma desnaturada. Ladero e colaboradores (2006) desenvolveram uma equação cinética que descreve essa situação, como pode ser observado na equação 15 e 16.

$$A = e^{-k_1 t} + \beta^* \frac{k_1}{k_2 - k_1} (e^{-k_1 t} - e^{-k_2 t}) \quad \text{Equação 16}$$

Onde A é a atividade remanescente relativa no tempo t, k_1 e k_2 são as constantes de inativação das etapas 1 e 2 respectivamente, e β^* é a atividade relativa do intermediário E^* .

Finalmente como um terceiro modelo, pode haver já inicialmente um equilíbrio entre duas formas ligeiramente diferentes da Sfβgli. Com isso, o resultado observado poderia ser uma

combinação entre as cinéticas de inativação térmica dessas duas formas. Ladero e colaboradores (2006) também descreveram esse processo, conforme as equações 17, 18 e 19.

$$A = \beta e^{-k_1 t} + (1 - \beta) (e^{-k_2 t}) \quad \text{Equação 19}$$

Onde β é um coeficiente que descreve a razão entre as atividades de ambas as formas da enzima presente em equilíbrio no início do experimento de inativação.

Apesar de serem hipóteses diferentes, equações 16 e 19 que descrevem esses diferentes modelos resultam em curvas semelhantes, tornando-se difícil distinguir qual dessas hipóteses descreve melhor o fenômeno observado para Sf β gli selvagem e mutantes. De fato, o ajuste dos dados experimentais para Sf β gli selvagem e mutantes nestes dois diferentes modelos rendeu coeficientes de correlação muito parecidos. Dessa forma, outros experimentos são necessários para distinguir entre essas hipóteses. Por outro lado, uma comparação com os experimentos de cinética de desnaturação acompanhados por *SyproOrange* (Figura 26) pode ser interessante neste sentido.

Apesar do comportamento de inativação não linear ter ocorrido para maioria das Sf β gli analisadas, nos experimentos da cinética de desnaturação térmica acompanhada por *SyproOrange* esse fenômeno não foi observado. Isso pode ter ocorrido porque nestes dois formatos de experimento a população de proteínas desnaturadas é estimada através de “sinais” diferentes, atividade enzimática *versus* ligação da sonda fluorescente. Deste modo, sob a perspectiva destes experimentos, desnaturação e inativação podem não ser necessariamente o mesmo fenômeno. Buscando ser mais claro, conforme a proteína desnatura, a fluorescência

aumenta devido à presença e ligação da sonda fluorescente *SyproOrange*, atingindo um patamar máximo de fluorescência. No tratamento e linearização dos dados, considera-se que neste patamar máximo, toda a população da proteína está desnaturada. No entanto, isso pode ser uma aproximação imprecisa, pois eventualmente com o aquecimento a proteína modifica sua estrutura, expõe porções apolares e liga o máximo possível de *SyproOrange* mesmo que não se desnature e inative completamente. Este estado intermediário poderia eventualmente ser cataliticamente ativo e seguir se desnaturando em uma velocidade menor do que aquela que levou a sua formação. Deste modo, se válida esta hipótese, no experimento acompanhado pela fluorescência de *SyproOrange* seria visualizada apenas a fase inicial mais rápida de formação deste intermediário, enquanto que ao monitorar a atividade seriam observadas duas fases, a formação e acúmulo do intermediário e em seguida o desaparecimento do mesmo.

Assumindo esta linha de raciocínio foram comparados na Tabela 5 os resultados de inativação (Figura 31) com os resultados de cinética de desnaturação de Sfβgli (Figura 26). Para tal foram empregadas apenas as constantes de velocidade da primeira etapa do processo de inativação térmica de Sfβgli selvagem e mutantes (Figura 31). Note que na tabela 5, as constantes de velocidade de inativação térmica foram multiplicadas por 2,3, para a comparação com as constantes de velocidade de desnaturação obtidas usando *SyproOrange*, pois a linearização dos dados de inativação foi feita usando logaritmo na base 10, e não o logaritmo natural, como no caso dos experimentos de desnaturação. Em linhas gerais, temos que os dados são semelhantes entre si, exibindo uma razão de 3 no máximo (Tabela 5), apenas os mutantes E265A, L389A e Y390A apresentam constantes com diferenças maiores que 4 vezes. De qualquer modo, a similaridade entre as constantes de velocidade de inativação e desnaturação das Sfβgli

selvagem e mutantes sugere que a primeira etapa da inativação térmica é o mesmo fenômeno acompanhado nos experimentos de desnaturação com *SyproOrange*, sendo então que a desnaturação de SfbgI segue o segundo modelo cinético discutido acima, ocorrendo em duas etapas com a formação de um intermediário entre elas.

Tabela 5. Constantes de velocidade de desnaturação e inativação térmica das SfbgI selvagem e mutantes.

<i>mutante</i>	<i>k_{desnaturação}</i> (<i>min</i> ⁻¹)	<i>k_{inativação}</i> (<i>min</i> ⁻¹)
WT	0,07±0,02	0,088±0,005
S247A	0,10±0,03	0,064±0,002*
N249A	0,4±0,1	0,23±0,02*
F251A	0,15±0,04	0,23±0,02*
F334A	0,29±0,08	0,152±0,004*
S358A	0,18±0,05	0,30±0,02*
K366A	0,35±0,09	0,23±0,02*
Y420A	0,6±0,2	0,53±0,04*
L100A	0,022±0,006	---
L350A	0,22±0,06	0,050±0,001*
D260A	0,18±0,05	0,34±0,01
E261A	0,033±0,008	0,118±0,007
M262A	0,035±0,009	0,111±0,005
A263F	0,20±0,05	0,13±0,01
A264F	Sem mudança	0,32±0,03
E265A	0,05±0,01	0,18±0,01
L335A	0,5±0,1	0,36±0,04
V336F	2,1±0,5	---
S337F	0,34±0,09	0,092±0,002
L389A	0,023±0,006	0,109±0,002
Y390A	0,04±0,01	0,14±0,02

*Resultados determinados por Ikegami (2013)

Tendo estabelecido o mecanismo de inativação de Sfβgli, analisando a figura 31, observa-se que dentre as mutações na vizinhança do resíduo central F251, as mutantes A264F e L335A apresentaram um perfil de inativação inicial com constante de velocidade maior do que a Sfβgli selvagem, sendo que as demais mutantes apresentaram valores de constante de inativação semelhantes à selvagem. De fato, a mutação A264F, que atinge um resíduo com cadeia lateral apontada para F251. Consistentemente, a mutação L335A também acelerou o processo de inativação, mas não é possível avaliar se esse efeito foi menor do que o da mutação V336F, pois esta mutante que apresenta atividade e concentração muito baixas. Em suma, mais uma vez mutações que afetam o resíduo central F251 tornaram o processo de inativação térmica mais rápido, observação coerente com a perda de estabilidade térmica já reportada acima.

2.4 Conclusões

Através dos resultados apresentados neste capítulo fica bastante evidente que as mutações que perturbam direta ou indiretamente os resíduos centrais da rede da estrutura da Sfβgli geram efeitos na sua estabilidade térmica.

Especificamente observa-se que mutações que substituem os resíduos centrais por alanina, rompendo grande parte de suas conexões com a vizinhança, resultam em diminuição da T_m da Sfβgli, e também promovem aumento de velocidade de desnaturação e de inativação. Por outro lado, mutações em resíduos que não perturbam resíduos centrais não tiveram, em sua maioria, efeitos significativos em nenhuma dessas propriedades.

Pode-se, então, supor que ao romper e/ou modificar as interações dos resíduos centrais com a sua vizinhança, a energia térmica, provavelmente, não se dissipa da mesma forma como na Sfβgli selvagem, pois resíduos centrais são fundamentais nesse processo de propagação da

energia térmica ao longo da estrutura (Szalay e Csermely, 2013). Dessa forma, a alteração na dissipação da energia térmica leva Sfβgli a assumir conformações menos estáveis, o que pode resultar na sua desnaturação (Leitner, 2008).

Além disso, as mutações na vizinhança do resíduo central F251 reforçam a hipótese de que os resíduos centrais são importantes na propagação térmica ao longo da estrutura, uma vez que as mutações em resíduos mais próximos e com a cadeia lateral apontada para o resíduo central geraram grandes efeitos na estabilidade térmica da Sfβgli.

Especificamente observou-se que a substituição do resíduo V336 (a 3,7 Å do resíduo central F251) por fenilalanina resultou em diminuição da T_m de 45 °C para 30 °C. Esta mesma mutação elevou em cerca de 30 vezes a velocidade de desnaturação de Sfβgli. Adicionalmente a mutação em outro resíduo muito próximo ao F251, a mutação A264F (a 3,0 Å do resíduo F251), fez com que a desnaturação térmica de Sfβgli não apresentasse uma transição característica (T_m), apresentando-se com múltiplas etapas, e não sendo do tipo “dois estados”. Isso indica que houve perda na cooperatividade do processo de desnaturação, um fenômeno já relatado por Shank e colaboradores (2010), que verificaram que resíduos centrais que conectam módulos dentro rede da estrutura de proteínas são fundamentais para manter a cooperatividade da desnaturação (Szalay e Csermely, 2013; Di Paola e Giuliani, 2015).

Em suma, os resultados apresentados nesse capítulo demonstram experimentalmente que, de fato, os resíduos centrais desempenham papéis fundamentais na propagação de energia térmica ao longo da estrutura proteica e que perturbações nos resíduos centrais podem levar a mudanças na termoestabilidade das proteínas.

3

CATÁLISE E AS REDES DE ESTRUTURA PROTEICA

3.1 Introdução

3.1.1 Glicosídeo Hidrolases – uma classe de enzimas com diversos papéis

Enzimas são biomoléculas fundamentais para a vida. Elas desempenham papéis essenciais nos organismos, catalisando as reações químicas para que elas ocorram com velocidades compatíveis com a vida (Berg *et al.*, 2012). Além desta importância biológica, enzimas têm sido utilizadas na catálise de muitos processos industriais (Kirk *et al.*, 2002; Singhanian *et al.*, 2016; Escamilla-Alvarado *et al.*, 2017). No entanto, para sua utilização em processos industriais, é necessário que a enzima possua características apropriadas, como a facilidade de produção e alta atividade, entre outras. Nesse sentido, diversos estudos têm buscado compreender as bases moleculares para a construção de enzimas adaptadas para processos industriais (Kirk *et al.*, 2002; Bornscheuer *et al.*, 2012; Yu e Huang, 2014).

As Glicosídeo Hidrolases (GH) são uma classe de enzimas largamente distribuída entre os seres vivos, sendo um grupo que catalisa reações de hidrólise envolvendo uma ampla gama de substratos e que reúne diversos tipos de enovelamento. Como existe uma grande diversidade de glicosídeos, oligo e polissacarídeos nos organismos, as GH estão associadas uma rica variedade

de funções. Devido à amplitude desse grupo, as GHs são divididas em famílias de acordo com sua similaridade de sequência. Como a similaridade está, muitas vezes, associada ao enovelamento e à função da proteína, a classificação das GHs reflete estas características (Lombard *et al.*, 2014). Dessa forma, até o presente momento, temos que as GHs estão divididas em 145 famílias. Uma das mais numerosas é a GH 1, na qual se encontram grande parte das beta-glicosidases.

3.1.2 Beta-glicosidases e sua importância

Beta-glicosidases (EC 3.2.1.21) catalisam a hidrólise de ligações glicosídicas de aril e alquil- β -glicosídeos ou ainda de oligocelodextrinas, liberando o resíduo β -D-glucosídico não-redutor (Cairns e Esen, 2010; Lombard *et al.*, 2014). Elas estão presentes em diversos organismos, como plantas, fungos, animais, arqueas, bactérias e até mesmo em vírus (Cairns e Esen, 2010; Antwerpen *et al.*, 2015). Nos organismos, as beta-glicosidases desempenham uma ampla variedade de funções, agindo no metabolismo de hidrólise da celulose, na hidrólise de glicolipídios, ativação de fito-hormônios, liberação de cianeto por plantas (cianogênese), digestão de glicosídeos, entre outras funções (Singh *et al.*, 2016). No entanto, apesar dessa variedade de funções, as beta-glicosidases pertencentes à GH1 são bastante conservadas, exibindo enovelamento do tipo barril (β/α)₈, mecanismo catalítico do tipo retentor da configuração do carbono anomérico e possuindo os mesmos resíduos catalíticos, dois ácidos glutâmicos (Glu) que desempenham papel de nucleófilo e de ácido/base (Marana, 2006).

Na figura 32 está representado o mecanismo catalítico das beta-glicosidases GH1, que possui basicamente duas etapas, sendo que uma delas é chamada de glicosilação e a outra,

desglicosilação. Na primeira etapa, a glicosilação, um dos glutamatos catalíticos faz um ataque nucleofílico no substrato, o que forma um intermediário onde a enzima fica covalentemente ligada à porção glicosídica do substrato. Simultaneamente o outro glutamato catalítico doa um próton ao produto que deixa o sítio ativo. Já na segunda etapa, a desglicosilação, uma molécula de água ativada pelo outro ácido glutâmico catalítico faz um ataque no intermediário, liberando o glicosídeo e a enzima livres (Rye e Withers, 2000). Alternativamente, o excesso de substrato ou ainda por modificações do solvente, essa segunda etapa pode sofrer modificações, sendo que uma nova molécula de substrato ou solvente é adicionada ao monossacarídeo do intermediário glicosil-enzima, gerando desta forma novos glicosídeos, um processo chamado de transglicosilação (Frutuoso e Marana, 2013). Além disso, Isorna e colaboradores discutem, em seu trabalho, que a tendência das beta-glicosidases por seguir uma ou outra rota também pode estar associada a um “canal de água”, por onde a molécula de água que participa na catálise acessa o sítio ativo (Isorna *et al.*, 2007; Frutuoso e Marana, 2013). Neste sentido, a polaridade deste canal, bem como o volume dos resíduos que o constituem fazem com que diferentes beta-glicosidases tenham diferentes tendências a hidrolisar ou transglicosilar (Frutuoso e Marana, 2013).

Além de grande importância para os organismos, as beta-glicosidases também tem despertado grande interesse quanto à aplicação em processos produtivos. Dentre esses processos, alguns estão relacionados com a atividade hidrolítica da enzima, como a produção de açúcares fermentáveis a partir de celulose, liberação de isoflavonas em sua forma ativa a partir de glucosídeos presentes na soja, detoxificação da mandioca através de hidrólise de glucosídeos cianogênicos, entre outros. Além disso, alguns processos se baseiam na “atividade de

transglicosilação” das beta-glicosidases, como a produção de oligossacarídeos e alquil-glicosídeos a partir de açúcares mais simples (Singhania *et al.*, 2016).

Figura 32. Mecanismo de catálise das beta-glicosidases GH1. Adaptado de Rye e Whitters (2000).

3.1.3 Bases moleculares do funcionamento das beta-glicosidases

Beta-glicosidases têm sido amplamente estudadas e o papel de diversos dos seus resíduos tem sido elucidado. Considerando a beta-glicosidase de *Spodoptera frugiperda* (Sfβgli), além dos dois resíduos catalíticos, E187 e E399, três resíduos têm se mostrado fundamentais para catálise, os resíduos R97 e Y331 responsáveis por modular o pK_a dos resíduos catalíticos e o

resíduo E329 que faz ligações de hidrogênio entre os dois resíduos catalíticos¹ Estes resíduos são amplamente conservados em beta-glicosidases, sugerindo que este papel é universal para as beta-glicosidases da família GH1. (Marana *et al.*, 2001; Marana *et al.*, 2003; Tamaki *et al.*, 2016).

O sítio ativo de beta-glicosidases é dividido em subsítios, que são regiões com tamanho suficiente para ligar uma unidade monossacarídica, conforme representado na figura 33 (Marana, 2006). O subsítio que liga a porção monossacarídica da extremidade não-redutora do substrato é chamado de subsítio de ligação do glicone, ou subsítio -1. Como diversas beta-glicosidases já foram cristalizadas com ligantes nessa região, as interações não-covalentes entre este subsítio e a porção glicone do substrato é bem conhecida para estas enzimas. Considerando que a Sf β gli interaja com a porção glicone de forma semelhante à β -glicosidase de *Neotermes koshunensis*, pois a posição dos resíduos no sítio ativo é muito semelhante entre as duas enzimas (Tamaki *et al.*, 2016), e também através de comparações com outras β -glicosidases GH1, podemos estimar as interações que ocorrem entre a enzima e o substrato na porção glicone. Desta forma, E399 e H142 parecem interagir com a hidroxila OH2 do substrato, Q39, H142 e W452 provavelmente fazem ligações de hidrogênio com a hidroxila OH3, Q39 e E451 fazem ligações de hidrogênio com OH4 e E451 interage desfavoravelmente com OH6 da porção do glicone. Além disso, um resíduo importante para a interação entre o glicone e o subsítio -1, é o W444, o qual é bastante conservado em beta-glicosidases e age como uma plataforma basal, fazendo interação do tipo *pi-stacking* com o glicone do substrato (Marana *et al.*, 2001; Marana, 2006; Tamaki *et al.*, 2016).

¹ A numeração dos resíduos considerada neste trabalho está baseada na sequência de aminoácidos da beta-glicosidase de *Spodoptera frugiperda* (Sf β gli) – ver Figura 5 e Figura 10. (Marana *et al.*, 2001)

Figura 33. Representação do sítio ativo de beta-glicosidases. Baseado Davies *et al.* (1997).

Além do subsítio -1, as beta-glicosidases também possuem subsítios de ligação ao aglicone. Considerando que o substrato em questão seja um alquil ou aril- β -glicosídeo, os subsítios de ligação ao aglicone ligam as porções alquila e arila do substrato. Alternativamente, ao considerar que o substrato em questão seja um oligossacarídeo, estes subsítios interagem com as porções monossacarídicas da extremidade redutora do oligossacarídeo. Cada porção da região de ligação do aglicone que tem tamanho suficiente para acomodar um monossacarídeo corresponde a um subsítio, sendo eles chamados de +1, +2, +3 e assim por diante, de acordo com a distância do ponto de clivagem do substrato (Figura 33) (Davies *et al.*, 1997). Beta-glicosidases possuem diferentes números de subsítios do aglicone, sendo que a beta-glicosidase de *Streptomyces ssp.* possui apenas o subsítio +1, enquanto que a beta-glicosidase de *Hordeum vulgare* possui subsítios +1, +2, +3 e +4 (Marana, 2006). Ao analisar a Sf β gli, temos que ela possui subsítios +1, +2 e +3 (Marana *et al.*, 2001).

Estudos da especificidade dos sítios de ligação ao aglicone e dados de estruturas cristalográficas têm mostrado que a interação entre estes subsítios e a porção aglicone do substrato, essencialmente, ocorre através de interações hidrofóbicas e, de forma pouco conservada, através de ligações de hidrogênio (Mendonca e Marana, 2008). Por exemplo, observa-se que a beta-glicosidase de *Zea mays*, que foi cristalizada com um ligante na sua região de ligação ao aglicone, possui basicamente interações do tipo hidrofóbicas, o que faz com que esta enzima aceite uma ampla variedade de substratos nessa porção. Por outro lado, a beta-glicosidase de *Sorghum bicolor*, que também foi cristalizada com a presença de ligante nessa região, possui diversas ligações de hidrogênio, o que faz com que ela tenha uma especificidade mais restritiva, só sendo ativa sobre o glicosídeo *dhuririn* (Czjzek *et al.*, 2000; Verdoucq *et al.*, 2004).

De modo geral, no subsítio +1 há a presença de uma plataforma basal, formada pela cadeia lateral do W351 no caso da Sf β gli, que é bastante conservado em beta-glicosidases GH1 (Marana, 2006). Essa plataforma faz interações hidrofóbicas com a porção aglicone do substrato, estabilizando-o. Desta forma, analisando a interação entre a beta-glicosidase Sf β gli e uma série de alquil- β -glicosídeos, observa-se que quanto mais carbonos na porção alquila, maior afinidade deste ligante pela enzima, sendo que esse aumento ocorre linearmente com o aumento do número de carbonos (Marana *et al.*, 2001). Adicionalmente à plataforma basal, um outro grupo de resíduos, chamado de *ceiling*, compõe o subsítio +1, sendo este bastante variável em beta-glicosidases. Assim, observa-se que os resíduos E190, E194, K201 e M453 formam o *ceiling* do subsítio +1 e contribuem para a especificidade do aglicone da Sf β gli (Mendonca e Marana, 2008).

Finalmente, Mendonça e Marana (2011) demonstraram através de mutagênese aleatória que curiosamente mesmo resíduos bem distantes do sítio ativo podem desempenhar papéis importantes na especificidade enzimática. Os autores observaram que os resíduos T35, Y345, P348, S358, S378, S424, F460 e R474, mesmo não fazendo parte do sítio ativo, são moduladores da especificidade enzimática pelo glicone do substrato. Desta forma, as mutações S378G, Y345C e S358F reduziram a especificidade por *p*-nitrofenil- β -fucopiranosídeo (*p*NPFuc) com relação ao *p*-nitrofenil- β -glucopiranosídeo (*p*NPGLu), enquanto que as mutações P348L, S424F, T35A, F460L e R474H aumentaram a especificidade do *p*NPFuc com relação a *p*NPGLu (Mendonça e Marana, 2011). Isso sugere que mutações fora do sítio ativo geram efeitos que são propagados e atingem o sítio ativo (Mendonça e Marana, 2011).

3.1.4 Redes de Estrutura Proteica e catálise enzimática

Para compreender e descrever a comunicação entre os resíduos da Sf β gli com o seu sítio ativo, Tamaki e colaboradores (2016) utilizaram uma abordagem de subdividir a estrutura de Sf β gli em camadas ao redor do sítio ativo. Neste estudo, a estrutura cristalográfica desta β -glicosidase (5CG0) foi usada para gerar um mapa de contatos entre resíduos, e com base nisso, considerando o sítio ativo como centro, os resíduos da Sf β gli foram distribuídos em camadas, de acordo com o número de conexões necessário para se atingir o sítio ativo. Observa-se neste trabalho que mutações em resíduos pertencentes até a camada 2 (dois contatos de distância de qualquer resíduo do sítio ativo) tiveram efeitos significativos na atividade enzimática. Sugere-se ainda que mutações em regiões mais próximas (em termos de número de contatos) do subsítio de interação com o glicone teriam efeitos deletérios, enquanto que

mutações em regiões mais próximas ao subsítio de ligação do aglicone teriam efeitos deletérios menos drásticos ou até benéficos à atividade enzimática de Sfβgli.

Já em uma abordagem distinta, outros trabalhos têm demonstrado que resíduos centrais da rede de estrutura proteica muitas vezes são importantes para a catálise enzimática. Amitai e colaboradores (2004), avaliando os resíduos centrais por *closeness* de 178 enzimas, observaram que estes resíduos frequentemente estão presentes no sítio ativo. No entanto, como nem todos os resíduos centrais da rede de estrutura proteica aparecem no sítio ativo, estes autores propuseram que para identificação do sítio ativo de enzimas desconhecidas deveríamos analisar não só a centralidade de seus resíduos, mas também a sua exposição ao solvente. De fato, esta combinação garantiu uma eficiência de 70% na detecção de sítio ativos usando um conjunto de enzimas previamente estudadas.

Dessa forma, baseando-se no exposto acima, uma pergunta em aberto é se mesmo os resíduos centrais da rede de estrutura proteica que são distantes do sítio ativo modulariam a atividade e especificidade enzimática. Neste sentido, o objetivo deste capítulo é avaliar se os resíduos centrais da rede de estrutura proteica distantes do sítio ativo desempenham papéis fundamentais para a catálise, podendo modular a especificidade pelo substrato mesmo que a distância. Para isso, a Sfβgli contendo mutações que perturbam os resíduos centrais, bem como Sfβgli contendo “mutações controle” que não perturbam os resíduos centrais, desenhadas segundo o capítulo 1, terão a sua atividade e especificidade pelo substrato avaliadas utilizando experimentos cinéticos e de efeito do pH na atividade enzimática.

3.2 Materiais e Métodos

3.2.1 Ensaio de atividade enzimática acompanhando formação de *p*-nitrofenol

Os ensaios de atividade da beta-glicosidase foram realizados utilizando 50 μL de amostra contendo a enzima e 50 μL do substrato *p*-nitrofenil- β -D-glucopiranosídeo (*p*NPGLu), *p*-nitrofenil- β -D-fucopiranosídeo (*p*NPFuc) e *p*-nitrofenil- β -D-galactopiranosídeo (*p*NPGal) em diversas concentrações, preparado em tampão (citrato-fosfato 100 $\text{mmol}\cdot\text{L}^{-1}$ pH 6,0, com exceção dos ensaios de pH ótimo, onde o pH foi modificado, ver item 3.2.4). A mistura de reação foi incubada a 30 °C por intervalos de tempo de 5, 10, 15 e 20 minutos e em seguida interrompida com 100 μL de solução de carbonato de sódio 0,5 $\text{mol}\cdot\text{L}^{-1}$. Nessas condições, o produto formado *p*-nitrofenol é convertido para a forma desprotonada, *p*-nitrofenolato, pois a solução de carbonato torna o meio alcalino. Com isso, o *p*-nitrofenolato foi detectado por meio da absorbância em 415 nm usando um leitor de placas Elx800 (Biotek®).

Ao final, as absorbâncias foram convertidas em mols de produto utilizando a curva padrão apresentada na Figura 34.

Figura 34. Curva-padrão de *p*-nitrofenolato

3.2.2 Ensaio de atividade enzimática utilizando celobiose como substrato

Os ensaios de atividade foram realizados utilizando 50 μL de amostra contendo a enzima e 50 μL celobiose em diversas concentrações preparada em tampão (citrato-fosfato 100 $\text{mmol}\cdot\text{L}^{-1}$ pH 6,0, com exceção dos ensaios de pH ótimo, onde o pH foi modificado, ver item 3.2.4). A mistura de reação foi incubada a 30 °C por intervalos de tempo de 5, 10, 15 e 20 minutos, e em seguida interrompida fervendo a amostra por 3 minutos. Após isso, a reação foi resfriada em gelo, e na sequência foram adicionados 200 μL de reagente para a detecção de glicose (PP-glicose, Analisa). Após incubação por 30 minutos a 37°C, a absorbância em 490 nm foi detectada usando um leitor de placas Elx800 (Biotek®).

Ao final, as absorbâncias foram convertidas em mols de produto (glicose) utilizando as curvas padrão apresentadas na Figura 35.

Figura 35. Curva padrão de glicose utilizando o reagente de detecção Glicose PP, da Analisa.

3.2.3 Ensaio de cinética enzimática

Para cálculo dos parâmetros cinéticos da S β gli, foram feitos ensaios de atividade enzimática utilizando diferentes concentrações de substratos (no mínimo dez). Como substrato foram utilizados *p*NPGLu, *p*NPFuc, *p*NPGal e celobiose. As curvas foram ajustadas segundo o modelo de Michaelis-Menten, de onde foram obtidos os parâmetros de V_{max} e K_m .

$$v = \frac{V_{max} \cdot [S]}{[S] + K_m} \quad \text{Equação 20}$$

O k_{cat} foi calculado a partir do V_{max} através da equação:

$$V_{max} = k_{cat} \cdot [E] \quad \text{Equação 21}$$

Para comparação dos valores de k_{cat} e k_{cat}/K_m das S β gli mutantes com a selvagem, foram consideradas significativas as diferenças de ao menos 3x. Isso porque esta diferença nos parâmetros significa mudanças energéticas maiores do que 3,0 kJ/mol. Com isso, mudanças de ao menos 3x nos parâmetros cinéticos podem indicar mudanças de ao menos uma ligação de hidrogênio entre enzima e o substrato no complexo ES e no complexo ativado ES ‡ .

3.2.4 Ensaio para determinação do efeito do pH na atividade enzimática

Para determinação do efeito do pH sobre a atividade de S β gli foram feitos ensaios utilizando *p*NPFuc 2mM em diferentes pHs. Como essa concentração de substrato está bem acima do K_m da S β gli selvagem e mutantes, considera-se que a velocidade inicial medida é

próxima à velocidade máxima ($V_{\max}$) da enzima. O substrato foi preparado em tampão citrato-fosfato 100 mM de pH 3,4 a 8,0, com variações de 0,2. Para evitar que outros fatores influenciem a atividade enzimática, todos os tampões foram preparados com a mesma força iônica, ajustando-a para 0,4 M com cloreto de sódio. Após o ensaio, gráficos do logaritmo da velocidade em função do pH foram plotados, e ajustados na equação 22 (Segel, 1993) utilizando a função de ajuste não linear do software Origin 8.5.1..

$$V = \frac{V_{\max}}{1 + \frac{10^{-pH}}{K_{ES1}} + \frac{K_{ES2}}{10^{-pH}}}$$

Equação 22

3.3 Resultados e Discussão

3.3.1 Efeitos na eficiência catalítica da Sfβgli promovidos por mutações na vizinhança de um resíduo central da rede estrutural

Foi anteriormente proposto que resíduos centrais desempenham papéis importantes na catálise enzimática (Amitai *et al.*, 2004; Del Sol *et al.*, 2006). Para testar experimentalmente esta hipótese, Ikegami (2013), em seu trabalho de tese, observou que mutações nos resíduos centrais da Sfβgli afetam, em alguns casos, os seus parâmetros cinéticos.

Contudo, como nestes resultados anteriormente obtidos não fica claro se há uma correlação entre perturbações nos resíduos centrais e efeitos na catálise, uma nova estratégia foi adotada nesta presente tese. Para isso, as Sfβgli D260A, E261A, M262F, A263F, A264F, E265A, L335A, V336F e S337F, contendo mutações na vizinhança do resíduo central F251, foram

planejadas e produzidas conforme descrito no capítulo 1. Brevemente estas mutações perturbam a vizinhança do resíduo central F251, e com isso, teoricamente esta perturbação poderia ser propagada para o sítio ativo, gerando mudanças na eficiência catalítica destas enzimas. Como os resíduos vizinhos do F251 estão dispostos em diferentes distâncias e posições relativas a este resíduo, uma correlação entre o posicionamento e o efeito na catálise torna-se possível, o que permite avaliar se, de fato, não há correlação entre centralidade dos resíduos e sua importância para a catálise. Estas enzimas mutantes foram então caracterizadas através de ensaios cinéticos usando os substratos *p*-nitrofenil- β -D-glucopiranosídeo (*p*NPGLu), *p*-nitrofenil- β -D-galactopiranosídeo (*p*NPGal), *p*-nitrofenil- β -D-fucopiranosídeo (*p*NPFuc) e celobiose.

Inicialmente devem ser discutidos os gráficos de $[P]$ *versus* t obtidos nos ensaios de atividade das S β gli A264F e V336F utilizando estes quatro substratos (Figura 36). Aparentemente, os gráficos de $[P]$ *versus* t apresentam-se lineares. No entanto, o intercepto no eixo y é inesperadamente alto, pois no início da reação não deveria haver produto. Controles feitos com a amostra contendo a enzima e também do substrato demonstraram que não se trata de absorbância inespecífica da amostra ou de produto contaminando o substrato. Com isso, uma possível explicação para este resultado é que as S β gli A264F e V336F podem estar se inativando durante o ensaio de atividade, que é feito a 30°C. De fato, a T_m da enzima V336F, (32 $\pm$ 3) °C, é muito próxima deste valor, e a mutante A264F nem sequer apresentou transição simples de forma nativa para desnaturada em função do aumento da temperatura (Capítulo 2). Desta forma, apesar da presença do substrato ter um papel importante de estabilização de enzimas por ligar preferencialmente a forma nativa da enzima, o que desloca o equilíbrio de desnaturação (Chaires, 2008), temos que esse efeito parece não ter sido suficiente para manter

as Sf β gli mutantes estáveis durante o período de 30 min utilizado no ensaio. Assim, a queda da concentração de enzima ativa geraria um gráfico do acúmulo do produto em função do tempo em curva tendendo a um patamar, do qual segmentos poderiam ser aproximados para retas com intercepto elevado. De qualquer modo, dado este comportamento e também a baixa atividade, não foi possível realizar a caracterização completa das Sf β gli A264F e V336F.

Figura 36. Ensaio de atividade enzimática das SfβglI A264F e V336F. Em A, B, C e D, ensaios da SfβglI A264F. Em E, F, G e H, ensaios da SfβglI V336F. Em A e E, foram utilizados como substrato pNPGlu 0,2 mM, em B e F, com substrato pNPFuc 0,2 mM, em C e G, com pNPGal 1,0 mM e em D e H, com celobiose 1 mmol·L⁻¹.

Já as demais S β gli mutantes (D260A, E261A, M262F, A263F, E265A, L335A e S337F) puderam ser caracterizadas normalmente. Nas figuras 37, 38, 39 e 40 estão os gráficos que relacionam a atividade das S β gli selvagem e destas mutantes com a concentração dos substratos *p*NPGLu, *p*NPFuc, *p*NPGal e celobiose, respectivamente.

Observa-se que para os gráficos obtidos com o substrato *p*NPGLu obteve-se comportamento Michaeliano para todas as enzimas analisadas (Figura 37). Por outro lado, ao analisar os resultados obtidos para o substrato *p*NPFuc, notou-se que havia uma diminuição da atividade em altas concentrações de substrato, o que não é previsto pelo modelo de Michaelis-Menten. Existem dois fenômenos encontrados em beta-glicosidases que podem resultar nesse desvio da equação de Michaelis-Menten. Um deles é que o excesso de substrato pode provocar inibição da enzima. Já o outro é a ocorrência de reações de transglicosilação do substrato (ver item 3.1.2), sendo que em altas concentrações de substrato a probabilidade de uma segunda molécula de substrato entrar no sítio ativo torna-se alta, favorecendo esse mecanismo. Se essa “rota de transglicosilação” for mais lenta do que a rota de hidrólise, há uma queda na atividade conforme foi observado. Deste modo, optou-se por utilizar os dados de velocidade obtidos para baixas concentrações de *p*NPFuc, os quais se ajustam ao modelo de Michaelis-Menten, permitindo a determinação dos parâmetros cinéticos para esse substrato (Figura 38).

Além disso, na figura 39, estão os gráficos que relacionam a atividade das S β gli em diferentes concentrações de *p*NPGal. Todas as enzimas ensaiadas com esse substrato apresentaram um comportamento Michaeliano. Por fim, a figura 40 apresenta os dados coletados para as S β gli selvagem e mutantes utilizando celobiose como substrato. Mais uma vez, os resultados com todas as enzimas foram curvas que se ajustam bem ao modelo de Michaelis-Menten.

Figura 37. Efeito da concentração do substrato pNPβGlu na velocidade inicial da reação catalisada pela enzima Sβgli selvagem e mutantes.

Figura 38. Efeito da concentração de substrato pNPβFuc na velocidade inicial da reação catalisada pela Sfβgli selvagem e mutantes.

Figura 39. Efeito da concentração de substrato $pNP\beta Gal$ na velocidade inicial da reação catalisada pela Sfgli selvagem e mutantes.

Figura 40. Efeito da concentração do substrato celobiose na velocidade inicial da reação catalisada pela Sfgli selvagem e mutantes.

A partir destes dados foram determinados os parâmetros cinéticos de cada uma das enzimas através do ajuste na equação de Michaelis-Menten. Na tabela 6 estão os valores de K_m das SfbgIi selvagem e mutantes utilizando diferentes substratos. Observa-se que nenhuma das mutações na vizinhança do resíduo central F251 causou mudanças significativas (de ao menos 3x) no K_m da SfbgIi, indicando que as mutações não causaram mudanças na afinidade desta enzima pelo substrato.

Tabela 6. Valores de K_m (em mmol·L⁻¹) para 4 tipos de substrato obtidos para a SfbgIi selvagem e mutantes

Enzima	Substrato			
	pNPGlu	pNPFuc	pNPGal	Celobiose
Selvagem	0,87 ± 0,04	0,81 ± 0,07	4,0 ± 0,2	5,0 ± 0,6
D260A	1,05 ± 0,06	0,82 ± 0,05	4,1 ± 0,3	7 ± 1
E261A	0,94 ± 0,01	0,80 ± 0,06	3,6 ± 0,2	4,5 ± 0,3
M262A	1,04 ± 0,07	1,4 ± 0,1	3,2 ± 0,4	4,5 ± 0,3
A263F	1,60 ± 0,06	0,64 ± 0,06	5,2 ± 0,4	6,8 ± 0,7
E265A	1,4 ± 0,1	1,5 ± 0,1	4,5 ± 0,4	7,8 ± 0,7
L335A	1,2 ± 0,1	0,61 ± 0,05	3,7 ± 0,2	5,8 ± 0,5
S337F	1,4 ± 0,1	0,45 ± 0,02	5,7 ± 0,5	7 ± 1

Por outro lado, nos valores de k_{cat} das SfbgIi selvagem e mutantes obtidos utilizando diferentes substratos indicados na Tabela 7, observa-se que as mutações A263F e S337F diminuíram o k_{cat} da enzima ao menos 3x para alguns dos substratos (células marcadas em verde), indicando que estas mutações provocam efeitos que prejudicam a catálise pela enzima. Nota-se que não há uma tendência quanto ao tipo de substrato cuja hidrólise é afetada, havendo substratos com hidroxila 4' equatorial (pNPGlu) e axial (pNPFuc) e com aglicone do tipo *p*-nitrofenil e glucosídeo (celobiose)

Tabela 7. k_{cat} (s^{-1}) das S β gli selvagem e mutantes para hidrólise de 4 tipos de substrato.

Enzima	Substrato			
	pNPGlu	pNPFuc	pNPGal	Celobiose
Selvagem	$2,56 \pm 0,02$	$5,8 \pm 0,2$	$0,322 \pm 0,006$	$2,14 \pm 0,08$
D260A	$1,63 \pm 0,03$	$2,95 \pm 0,09$	$0,223 \pm 0,004$	$1,45 \pm 0,07$
E261A	$2,10 \pm 0,06$	$5,3 \pm 0,2$	$0,310 \pm 0,006$	$1,89 \pm 0,04$
M262A	$2,64 \pm 0,05$	$7,9 \pm 0,5$	$0,32 \pm 0,01$	$2,13 \pm 0,04$
A263F	$0,806 \pm 0,008$	$1,35 \pm 0,05$	$0,194 \pm 0,007$	$0,72 \pm 0,04$
E265A	$1,86 \pm 0,04$	$6,4 \pm 0,5$	$0,278 \pm 0,008$	$1,74 \pm 0,07$
L335A	$1,57 \pm 0,05$	$3,4 \pm 0,1$	$0,247 \pm 0,005$	$1,07 \pm 0,03$
S337F	$0,96 \pm 0,02$	$1,24 \pm 0,01$	$0,147 \pm 0,005$	$0,50 \pm 0,03$

Finalmente, foi calculada a razão entre esses parâmetros, k_{cat}/K_m , denominada “eficiência catalítica”, pois reflete a especificidade da enzima considerando a sua afinidade pelo substrato e a velocidade da catálise. Esse parâmetro para cada uma das enzimas está representado na tabela 8. Observa-se que, de forma semelhante ao analisar somente o k_{cat} , duas mutações na vizinhança do F251, as mutações A263F e S337F, provocaram mudanças no k_{cat}/K_m , diminuindo-o em pelo menos 3x do valor da S β gli selvagem para alguns dos substratos (células em verde). Novamente não há uma tendência quanto ao tipo de substrato cuja hidrólise é afetada.

Tabela 8. k_{cat}/K_m ($mM^{-1}s^{-1}$) das S β gli selvagem e mutantes para 4 tipos de substrato.

Enzima	Substrato			
	pNPGlu	pNPFuc	pNPGal	Celobiose
Selvagem	$2,9 \pm 0,1$	$7,1 \pm 0,7$	$0,080 \pm 0,004$	$0,42 \pm 0,05$
D260A	$1,5 \pm 0,1$	$3,6 \pm 0,2$	$0,054 \pm 0,004$	$0,21 \pm 0,03$
E261A	$2,23 \pm 0,07$	$6,6 \pm 0,5$	$0,086 \pm 0,005$	$0,42 \pm 0,03$
M262A	$2,5 \pm 0,2$	$5,7 \pm 0,5$	$0,10 \pm 0,01$	$0,47 \pm 0,03$
A263F	$0,50 \pm 0,05$	$2,1 \pm 0,2$	$0,037 \pm 0,003$	$0,11 \pm 0,01$
E265A	$1,3 \pm 0,1$	$4,3 \pm 0,4$	$0,062 \pm 0,006$	$0,22 \pm 0,02$
L335A	$1,3 \pm 0,1$	$5,6 \pm 0,5$	$0,067 \pm 0,004$	$0,18 \pm 0,02$
S337F	$0,68 \pm 0,05$	$2,7 \pm 0,1$	$0,026 \pm 0,002$	$0,07 \pm 0,01$

3.3.2 Cálculo de mudança energética de catálise provocada pelas mutações na vizinhança de um resíduo central da rede estrutural

Para compararmos diretamente os valores dos parâmetros cinéticos das SfbgI contendo mutações na vizinhança do resíduo F251 (Tabelas 6, 7 e 8), estes foram transformados em valores de variação de energia livre (ΔG) dos processos aos quais estes parâmetros estão associados. Adicionalmente, para objetividade da comparação assumiu-se que um efeito mutacional relevante deveria causar uma mudança de energia livre (ΔG) ao menos igual a uma ligação de hidrogênio (3,0 kJ/mol) (Fersht *et al.*, 1999).

Os cálculos efetuados basearam-se no esquema da variação de energia livre em uma reação catalisada por uma enzima (Figura 41). Neste esquema, é possível observar que a ligação entre enzima e substrato está associada a uma mudança energética, representada por ΔG_b . Além disso, o valor de energia da “barreira” entre ES e ES[‡] é representada no esquema por $\Delta G_{cat}^{\ddagger}$. Já o valor da “barreira energética” com relação ao estado inicial (E + S) é representado no esquema como $\Delta G^{\ddagger}$, sendo que esse valor nada mais é do que a soma dos dois valores citados anteriormente. Esses valores podem ser calculados utilizando as equações 21, 22 e 23, que foram extraídas de Warshel e Florián (1998).

$$\Delta G_b \cong RT \ln \frac{1}{K_m} \quad \text{Equação 23}$$

$$\Delta G_{cat}^{\ddagger} = RT \ln \left(\frac{k_{cat}}{A} \right) \quad \text{Equação 24}$$

$$\Delta G^{\ddagger} = RT \ln \left(\frac{k_{cat} / A}{K_m} \right) \quad \text{Equação 25}$$

Onde R é a constante dos gases, T a temperatura absoluta e A um fator pré-exponencial igual a $k_B T/h$, sendo k_B a constante de Boltzmann e h a constante de Planck.

Figura 41. Perfil energético de uma reação catalisada por enzima. Adaptado de Warshel e Florián (1998).

Com base nisso, podemos calcular a diferença entre os ΔG comparando a SfbgI selvagem com cada um dos mutantes, considerando cada um dos substratos testados. Estes $\Delta\Delta G_b$, $\Delta\Delta G_{cat}^\ddagger$ e $\Delta\Delta G^\ddagger$ avaliam a mudança energética no perfil da reação catalisada enzimaticamente causada pela mutação em cada um desses parâmetros. Considerando um esquema representando o perfil energético da catálise de uma enzima selvagem (linha preta) e da mesma enzima após uma mutação (linha vermelha) (Figura 42), vemos que o $\Delta\Delta G_b$ é a mudança energética provocada pela mutação na ligação da enzima com o substrato, $\Delta\Delta G^\ddagger$ é a mudança provocada pela mutação na energia de ativação da catálise e que o $\Delta\Delta G_{cat}^\ddagger$, por esse esquema, é a mudança resultante da mutação na altura da barreira para atingir o estado de transição “normalizando” a energia do

estado ES da enzima selvagem e da enzima mutante para um mesmo patamar (inserto da Figura 42).

Figura 42. Efeito de uma mutação no perfil energético da reação catalisada por uma enzima. No inserto da figura, temos que o $\Delta\Delta G^{\ddagger}_{cat}$ é a energia da barreira do estado de transição “normalizando” as energias do estado ES da enzima selvagem e mutante, ou seja, é a diferença de energia de ativação considerando que o complexo ES possui a mesma energia entre as enzimas, e como isso não é necessariamente verdadeiro, é como se normalizássemos os dois estados ES para o mesmo nível de energia. Já a figura abaixo mostra que o $\Delta\Delta G^{\ddagger}$ é a diferença de energia entre os estados $ES^{\ddagger}$. Baseado em Warshel e Florian (1998).

Dessa forma, os valores de $\Delta\Delta G_b$, o $\Delta\Delta G^{\ddagger}_{cat}$ e o $\Delta\Delta G^{\ddagger}$ foram calculados utilizando as equações 24, 25 e 26, respectivamente. Os resultados obtidos estão representados na figura 43.

$$\Delta\Delta G_b = RT \ln \left(\frac{K_m(mut)}{K_m(selv)} \right) \quad \text{Equação 26}$$

$$\Delta\Delta G^{\ddagger}_{cat} = RT \ln \left(\frac{k_{cat}(selv)}{k_{cat}(mut)} \right) \quad \text{Equação 27}$$

$$\Delta\Delta G^\ddagger = RT \ln \left(\frac{\frac{k_{cat}(selv)}{K_m(selv)}}{\frac{k_{cat}(mut)}{K_m(mut)}} \right) \quad \text{Equação 28}$$

Onde $k_{cat}(selv)$ e $K_m(selv)$ correspondem ao k_{cat} e o K_m da Sfβgli selvagem, e $k_{cat}(mut)$ e $K_m(mut)$ correspondem ao k_{cat} e K_m da Sfβgli mutante.

Na figura 43A é possível observar que as mutações tiveram pouco efeito na energia de ligação entre enzima e substrato (ΔG_b). Vale repetir que o parâmetro de significância assumido foi a energia típica de ligações de hidrogênio (a partir de 3,0 kJ/mol) (Fersht, 1999). Com isso nota-se que as mutações introduzidas em Sfβgli não parecem ter tido efeito suficientemente a ponto de mudar o saldo de ligações de hidrogênio no complexo enzima-substrato.

Já ao observar as mudanças na energia de ativação (Figuras 43B e 43C), nota-se que ocorreram aumentos significativos da barreira energética $\Delta G_{cat}^\ddagger$ e $\Delta G^\ddagger$, dificultando a catálise. De fato, duas mutações na vizinhança do resíduo central F251, as mutações A263F e S337F, provocaram aumentos maiores do que 3,0 kJ/mol para hidrólise de *p*NPGLu, *p*NPFCu e celobiose nessas barreiras. Dentre todos os resíduos mutados na vizinhança de F251 que puderam ser caracterizados quanto à atividade enzimática, S337 e A263 são aqueles que mais próximo se posicionam do resíduo central F251 (Figura 23, capítulo 2). Estas mutações também reduziram significativamente a T_m de Sfβgli (Capítulo 2).

Figura 43. Efeito das mutações no perfil energético da reação catalisada por SfbgI. A, Efeito das mutações na energia de ligação do substrato (ΔG_b). B, Efeito da mutação na energia de ativação $\Delta G_{cat}^{\ddagger}$. C, Efeito da mutação na energia de ativação total $\Delta G^{\ddagger}$.

Em suma, nota-se que o principal componente nas mudanças energéticas no perfil da reação catalisada causadas pelas mutações na vizinhança do resíduo central F251 da rede estrutural de Sfβgli é o aumento na energia de ativação da reação ($\Delta G^{\ddagger}_{\text{cat}}$ e $\Delta G^{\ddagger}$), sendo que o efeito na energia de ligação do substrato foi pouco significativo. Esta tendência está em linha com a proposta geral que enzimas estabilizam preferencialmente o estado de transição da reação (Fersht, 1999). Como estas mutações estão distantes do sítio ativo em pelo menos 20 Å, estes efeitos mutacionais podem ter ocorrido devido a uma modificação local que propagou e atingiu globalmente a estrutura de Sfβgli, possivelmente resultando na estabilização de um estado conformacional que exhibe alterações estruturais no sítio ativo que mesmo sendo sutis de alguma forma prejudicam a catálise.

3.3.3 Mudanças de especificidade provocadas pelas mutações na vizinhança de um resíduo central da rede estrutural

Como visto acima as mutações na vizinhança do resíduo central F251 provocaram mudanças no perfil energético da reação catalisada por Sfβgli, o que possivelmente ocorre devido a mudanças no sítio ativo desta enzima. Essas mudanças podem atingir de maneira igual a estabilização do estado de transição de todos os substratos, ou agir mais seletivamente. Quando este último caso ocorre, temos uma mudança na especificidade.

Como ponto de partida para avaliar se houve mudanças de especificidade das Sfβgli mutantes, foram comparados os valores de k_{cat}/K_m relativos para as enzimas selvagem e mutantes sobre os 4 substratos ilustrados na figura 44. Este parâmetro é calculado dividindo-se o k_{cat}/K_m para os diferentes substratos de uma mesma enzima pelo k_{cat}/K_m mais alto encontrado

em um dos seus substratos. Dessa forma é possível normalizar os valores, desconsiderando mudanças de k_{cat}/K_m que atingem igualmente todos os substratos.

Figura 44. Estrutura dos substratos utilizados para a análise da especificidade das Sfβgli selvagem e mutantes.

Os valores obtidos para a Sfβgli selvagem e mutantes estão representados na figura 45. Através desta análise, não é possível observar mudanças de especificidade provocadas pelas mutações, sendo que a especificidade relativa da enzima selvagem, maior preferência por pNPFuc, seguido por pNPGlu, celobiose e pNPGal, parece conservada nas mutantes.

Figura 45. Valores de k_{cat}/K_m relativo da SfbgI selvagem e mutantes. wt, selvagem.

Uma forma mais detalhada de avaliar o efeito das mutações na especificidade é comparar dois sítios ativos utilizando a análise de *Linear Free Energy Relationship* - LFER (Warshel *et al.*, 1994). Para isso, é calculada a diferença na energia de ativação aparente entre pares de substratos para a mesma enzima, segundo a equação 27.

$$\Delta\Delta G^\ddagger = RT \ln \left(\frac{\frac{k_{cat}}{K_m}(\text{substrato 1})}{\frac{k_{cat}}{K_m}(\text{substrato 2})} \right) \quad \text{Equação 29}$$

Esse valor é calculado para todas as duplas possíveis de substratos para uma mesma enzima. Após isso, gráficos são construídos relacionando valores obtidos para a SfbgI selvagem com valores obtidos para uma mutante. Dessa forma, é possível avaliar se as mutações interferem diferencialmente no $\Delta G^\ddagger$. Se o resultado desse gráfico for uma reta com inclinação 1, o sítio ativo da SfbgI selvagem e da SfbgI mutante são iguais, ou seja, estabilizam $S^\ddagger$ (estado de

transição do substrato) do mesmo modo, logo não houve mudanças de especificidade. No entanto, se o resultado for uma reta com inclinação diferente de 1, temos que a mutação provocou uma mudança na especificidade da Sf β gli. Por fim, caso o resultado não seja uma reta, temos que houve uma mudança no mecanismo de catálise da Sf β gli mutante.

Dessa forma, foram feitos os cálculos de $\Delta\Delta G^\ddagger$ para cada dupla de substrato hidrolisados pelas Sf β gli selvagem e mutantes (Tabela 9). Valores negativos indicam que o segundo substrato do par é hidrolisado mais rapidamente e o inverso é válido para valores positivos. Observa-se que para Sf β gli selvagem e todas as mutantes, a reação utilizando um substrato com a hidroxila 4' na posição equatorial (como em glicose) é mais favorecida, assim como a remoção da hidroxila 6' (como em fucose) também torna o substrato mais favorável à hidrólise. Além disso, as Sf β gli têm preferência maior por substratos com grupos *p*-nitrofenil do que glicosídico na região do aglicone. Estas tendências estão de acordo com as descritas por Marana *et al.* (2001).

Tabela 9. $\Delta\Delta G^\ddagger$ (kJ/mol) entre diferentes substratos para as Sf β gli selvagem e mutantes.

Enzima	$\Delta\Delta G^\ddagger$ (glu-gal)	$\Delta\Delta G^\ddagger$ (gal-fuc)	$\Delta\Delta G^\ddagger$ (fuc-glu)	$\Delta\Delta G^\ddagger$ (celo-glu)	$\Delta\Delta G^\ddagger$ (celo-fuc)	$\Delta\Delta G^\ddagger$ (celo-gal)
Selvagem	9,08	-11,32	2,24	-4,85	-7,09	4,22
D260A	8,46	-10,58	2,12	-5,07	-7,19	3,38
E261A	8,38	-11,12	2,74	-4,20	-6,94	4,18
M262A	8,17	-10,20	2,02	-4,25	-6,27	3,92
A263F	6,56	-10,19	3,63	-3,91	-7,53	2,65
E265A	7,72	-10,66	2,94	-4,50	-7,44	3,22
L335A	7,49	-11,14	3,60	-4,95	-8,59	2,54
S337F	8,22	-11,74	3,52	-5,69	-9,21	2,53

Com os dados da tabela 9 foram feitos os plots LFER comparando as Sf β gli selvagem e mutantes (Figura 46).

Figura 46. Análise LFER comparando as Sβgli selvagem e mutantes.

Na figura 46 observa-se que as LFER para todas as Sβgli mutantes tiveram inclinação próxima a 1,0 quando comparadas com a enzima selvagem, revelando que estes mutantes ligam e estabilizam o estado de transição do substrato do mesmo modo que a enzima selvagem. Isso indica que as mutações na vizinhança do resíduo F251, apesar de perturbá-lo e terem seu efeito propagado alterando a catálise (ver item anterior), não modificaram a especificidade da enzima pelo substrato. Dessa forma, as perturbações foram propagadas e atingiram o sítio ativo, mas

resultaram em mudanças equivalentes na interação com o estado de transição dos quatro substratos testados.

3.3.4 Efeito das mutações na vizinhança de um resíduo central no pH ótimo da S β gli

Uma segunda opção para avaliar se as mutações próximas ao resíduo central F251 provocam perturbações que são propagadas até o sítio ativo é a comparação dos pK_a s dos resíduos catalíticos, E187 e E399, entre as S β gli selvagem e mutantes. De fato, mudanças estruturais no sítio ativo poderiam alterar o ambiente e interações não-covalentes dos resíduos catalíticos com outros resíduos, o que se refletiria em mudanças nos pK_a s dos resíduos catalíticos e no pH ótimo da enzima.

Para determinar os valores de pK_a dos resíduos catalíticos das enzimas selvagem e mutantes foram feitos ensaios de atividade utilizando *p*NPFuc em concentração 2,5 vezes maior do que o K_m (Figura 47). Nessas condições, a velocidade inicial medida torna-se próxima à velocidade máxima. Assim, variações da atividade refletem diretamente o efeito do pH sobre o parâmetro cinético V_{max} , o qual é diretamente proporcional ao k_{cat} . Além disso, todos os ensaios foram realizados na mesma força iônica, 0,4 M, para distinguir os efeitos do pH dos efeitos da força iônica, que se altera quando são usados diferentes sistemas tamponantes.

Figura 47. Comparação do efeito do pH na atividade enzimática das SfbgI selvagem e mutantes. As esferas pretas são os dados coletados para os mutantes, enquanto que os losangos vermelhos são os dados coletados para a enzima selvagem.

Finalmente as curvas foram ajustadas na Equação 22 (ver materiais e métodos) para obtenção dos pK_a s dos resíduos catalíticos (Tabela 6). Na Figura 47, pode ser observado que os dados para valores de pH muito baixos não foram considerados no ajuste porque a perda de atividade nesta faixa pode estar associada a outros fenômenos além da protonação dos resíduos catalíticos, como por exemplo a instabilidade estrutural.

Tabela 10. Valores de pK_a dos resíduos catalíticos e pH ótimo das Sfβgli selvagem e mutantes.

<i>Enzima</i>	<i>pK_{a1}</i>	<i>pK_{a2}</i>	<i>pH ótimo</i>
Selvagem	$5,6 \pm 0,4$	$7,5 \pm 0,5$	$6,5 \pm 0,6$
D260A	$6,0 \pm 0,3$	$7,3 \pm 0,4$	$6,7 \pm 0,5$
E261A	$5,8 \pm 0,2$	$7,4 \pm 0,3$	$6,6 \pm 0,3$
M262A	$5,8 \pm 0,3$	$7,3 \pm 0,3$	$6,6 \pm 0,4$
A263F	$5,4 \pm 0,4$	$7,2 \pm 0,4$	$6,3 \pm 0,6$
E265A	$6,3 \pm 0,5$	$7,3 \pm 0,6$	$6,8 \pm 0,8$
L335A	$6,0 \pm 0,5$	$7,5 \pm 0,6$	$6,7 \pm 0,8$
S337F	$5,0 \pm 0,2$	$7,5 \pm 0,2$	$6,2 \pm 0,3$

pK_{a1} corresponde ao nucleófilo catalítico E399 e pK_{a2} ao ácido catalítico E187.

Observa-se pela tabela 10 que o valor de pK_{a2} , que correspondem ao ácido catalítico E187, não é modificado por nenhuma das mutações. O valor de pH ótimo da enzima também não é significativamente afetado por nenhuma das mutações. No entanto, o valor de pK_{a1} , que corresponde ao nucleófilo E399, parece ter sido afetado pela mutação S337F (Tabela 10). Esta comparação torna-se mais clara na sobreposição entre as curvas de efeito do pH sobre a atividade da Sfβgli selvagem e mutantes (Figura 47). Percebe-se que o ramo da curva abaixo do pH ótimo, que é preponderantemente determinada pelo pK_{a1} , mostra-se significativamente diferente entre a Sfβgli selvagem e o mutante S337F. O pK_a do nucleófilo catalítico E399 é modulado por interações com dois outros resíduos centrais de Sfβgli, R97 e Y331 (Marana *et al.*, 2003), portanto E399 está relativamente próximo em termos de contatos do restante da enzima e também de outros resíduos centrais como F251. Destaca-se que S337F foi uma mutação que

aumentou significativamente o $\Delta\Delta G^\ddagger$ e a T_m de Sf β gli, sendo um dos resíduos mais próximos ao central F251 (Figura 23, capítulo 2).

Estes resultados indicam que, de uma forma geral, as mutações que perturbam o resíduo central F251 não afetaram os pK_a s dos resíduos catalíticos, sendo que somente uma das mutações mais próximas ao resíduo central, S337F, provocou uma leve redução no pK_{a1} , sugerindo que pode ter modificado o ambiente e interações do resíduo catalítico E399, apesar de se localizar distante dele.

3.3.5 Correlação entre termoestabilidade e mudanças no sítio ativo

Anteriormente no capítulo 2 foram descritos os efeitos das mutações na vizinhança do resíduo F251 sobre a termoestabilidade de Sf β gli, indicando que mutações em resíduos próximos do resíduo central F251 causaram maiores reduções na T_m da enzima. Já neste capítulo, foram apresentados resultados mostrando que mutações próximas ao F251, como S337F e A263F, também afetam o sítio ativo desta enzima, modificando os parâmetros cinéticos (especialmente o k_{cat} e k_{cat}/K_m) e elevando a barreira de ativação da reação catalisada por Sf β gli ($\Delta G^\ddagger$ e $\Delta G^\ddagger_{cat}$), mas sem afetar significativamente a especificidade desta enzima. Naturalmente é interessante observar se os efeitos dos parâmetros catalíticos estão correlacionados aos efeitos de termoestabilidade observados. A figura 48 mostra um gráfico comparando a T_m das Sf β gli selvagem e mutantes e os seus k_{cat}/K_m para o substrato pNPGlu.

Figura 48. Relação entre a termoestabilidade e a eficiência catalítica da SfbgI selvagem e mutantes que perturbam o resíduo central F251. gluco corresponde à pNPGlu.

Observa-se uma grande similaridade entre o perfil dos efeitos mutacionais sobre estas duas propriedades quando analisamos as mutações na vizinhança do resíduo central F251. Deste modo, termoestabilidade e atividade catalítica flutuam conjuntamente, com um perfil muito semelhante quando comparadas com a enzima selvagem.

Contrariamente a esses resultados, Shoichet e colaboradores (1995) sugeriram que a eficiência catalítica é inversamente proporcional a termoestabilidade da proteína. Isso porque supõem que a eficiência catalítica é associada a uma maior liberdade dinâmica da proteína, enquanto a termoestabilidade estaria associada a uma maior rigidez e compactação da estrutura. Dessa forma, o ganho de mobilidade diminuiria a termoestabilidade, mas aumentaria a eficiência catalítica da proteína. Por outro lado, Lehmann *et al.* (2000) discutem que termoestabilidade e eficiência catalítica parecem ser propriedades independentes e apontam resultados onde o aumento da termoestabilidade resultou em aumento da eficiência catalítica.

No entanto, é curioso destacar que os resultados aqui apresentados sugerem que termoestabilidade e atividade catalítica parecem correlacionadas, mas de uma forma diferente da proposta por Shoichet *et al.* (1995).

Além da comparação traçada na figura 48, pondera-se que do mesmo modo que as mutações que provocaram maiores efeitos na termoestabilidade são justamente aquelas mais próximas ao resíduo central F251 (V336F e A264F, capítulo 2), observa-se que as mutações S337F e A263F, também próximas do F251, afetaram a catálise, principalmente através da diminuição do k_{cat} e k_{cat}/K_m e consequente elevação da barreira de ativação da reação catalisada por SfbgI ($\Delta G^\ddagger$ e $\Delta G^\ddagger_{cat}$) (Figura 23, Tabela 7 e Tabela 8). Uma destas mutações, S337F, também causou uma mudança no pK_a do nucleófilo catalítico E399 (Figura 47 e Tabela 10).

Assim, estas coincidências reforçam a hipótese de que os efeitos observados ocorrem, de fato, através do resíduo central F251, cuja perturbação é eficientemente propagada, atingindo o sítio ativo. Em tese, a rede de interações da estrutura proteica é alterada pelas mutações na vizinhança do resíduo central, o que resulta em uma mudança na termoestabilidade. Adicionalmente como estas modificações estruturais são propagadas pela rede de interações para toda a proteína, elas também chegam ao sítio ativo (Figura 49), o qual é afetado e perde capacidade catalítica.

Figura 49. Esquema representando a propagação do efeito das mutações na vizinhança de um resíduo central (*hub*) pela proteína chegando ao sítio-ativo.

3.4 Conclusões

Notou-se que as mutações D260A, E261A, M262F, A263F, A264F, E265A, L335A, V336F e S337F introduzidas na vizinhança do resíduo central F251 em Sf β gli não causaram efeito no saldo de ligações de hidrogênio no complexo enzima-substrato (ES), não havendo mudança significativa do K_m para quatro substratos analisados (pNPGlu, pNPGal, pNPFuc e celobiose). Já ao analisar as mudanças na energia de ativação, nota-se duas mutações, A263F e S337F, que provocaram aumento significativo ($>3,0$ kJ/mol) da barreira energética $\Delta G_{cat}^\ddagger$ e $\Delta G^\ddagger$, dificultando a catálise. Dentre todos os resíduos mutados na vizinhança de F251 que puderam ser caracterizados quanto à atividade enzimática, S337 e A263 são aqueles que mais próximo se posicionam do resíduo central F251. Apesar destes efeitos na energia de ativação, a análise de LFER não mostrou alterações na especificidade pelo substrato em nenhuma das Sf β gli mutantes.

Adicionalmente a determinação do efeito do pH na atividade enzimática destas Sf β gli mutantes mostrou que apenas a mutação S337F causou redução no pK_a do nucleófilo catalítico E399, apesar desta mutação estar a cerca de 20 Å do sítio ativo desta enzima.

Por fim, observou-se uma grande similaridade entre os efeitos sobre a T_m e o k_{cat}/K_m quando analisamos as mutações D260A, E261A, M262F, A263F, E265A, L335A e S337F na vizinhança do resíduo central F251. De fato, a termoestabilidade e atividade catalítica da Sf β gli flutuam conjuntamente, com um perfil muito semelhante quando comparadas com a enzima selvagem.

Em conclusão, os resultados obtidos neste capítulo 3 reforçam a hipótese já levantada no capítulo 2 de que os efeitos decorrentes da perturbação do resíduo central F251 são eficientemente propagados pela rede de interações para toda a proteína, inclusive, como agora evidenciado, atingindo e alterando o sítio ativo de Sf β gli.

4

ESTUDO EXPERIMENTAL DO ESTADO OLIGOMÉRICO DA SF β GLI E PERSPECTIVAS UTILIZANDO A ANÁLISE POR REDES DE ESTRUTURA PROTEICA

4.1 Introdução

4.1.1 Importância do estado oligomérico de enzimas

Durante muitos anos, estudos têm elucidado estrutura de proteínas formando dímeros, tetrâmeros e outros tipos de arranjo (Janin *et al.*, 2008). Nestes estudos, as principais propriedades que têm sido relacionadas com estados oligoméricos de enzimas são a regulação e cooperatividade. Neste sentido, a formação de arranjos oligoméricos permite uma regulação da função fisiológica dessas proteínas, o que é de extrema importância nos organismos, pois através dessa regulação, as reações biológicas podem ser ajustadas de acordo com o estado celular (Janin *et al.*, 2008). Apesar disso, os princípios dessa formação oligomérica nem sempre são bem compreendidos.

4.1.2 Técnicas para caracterização de estados oligoméricos de proteínas

Diversas técnicas têm sido utilizadas para caracterização de estados oligoméricos de proteínas. O uso de um agente formador de ligações covalentes cruzadas e subsequente análise

por eletroforese SDS-PAGE, cromatografia de gel-filtração, entre outras técnicas, têm sido tradicionalmente usadas em Bioquímica. No entanto, estas técnicas são mais qualitativas, trazendo poucas informações sobre as quantidades reais das espécies oligoméricas em solução. Nesse contexto, a técnica de Ressonância Magnética Nuclear (RMN) vem ganhando bastante espaço (Zuiderweg, 2002). No entanto, a técnica de RMN ainda apresenta como limitante o tamanho do complexo (Wuthrich, 1990; Zuiderweg, 2002).

Com isso, as técnicas de espalhamento têm sido apontadas como bastante promissoras para a caracterização oligomérica de proteínas. A cristalografia de raios-X, muito utilizada para a caracterização de estrutura terciária de proteínas, algumas vezes também é utilizada para caracterizar sua estrutura quaternária (Banerjee *et al.*, 1994). No entanto, como essa análise é feita com a proteína na forma cristalina, muitas vezes a estrutura quaternária encontrada é diferente da estrutura em solução (Levy, 2007). Para contornar esta dificuldade, alguns estudos têm utilizados programas de predição de estrutura quaternária associados aos resultados de cristalografia (Krissinel e Henrick, 2007; Levy, 2007).

Além disso, uma forma empírica de avaliar o estado oligomérico de proteínas em solução é o uso de espalhamento a baixos ângulos (SAS), especialmente através das técnicas de Espalhamento de Raios-X a baixos ângulos (SAXS) e Espalhamento de Nêutrons a baixos ângulos (SANS) (Svergun, 2007). SAXS é uma técnica análoga à difração de raios-X feita na cristalografia, sendo que sua principal diferença é que a amostra é analisada em solução aquosa. Devido às partículas se encontrarem em solução aquosa, elas ficam randomicamente orientadas, e com isso há uma perda de informação quando comparado à cristalografia (Figura 50) (Svergun, 2007; Mertens e Svergun, 2010). Apesar disso, como a análise é feita em solução, a estrutura

quaternária pode ser mais bem avaliada, determinando-se inclusive as condições necessárias para mantê-la, como força iônica, pH, etc. SANS é uma técnica análoga a SAXS, mas que utiliza um feixe de nêutrons térmicos ao invés de raios-X (Svergun, 2007). Como nêutrons são partículas pesadas, eles colidem com o núcleo dos átomos da amostra, resultando em espalhamento. A principal diferença entre as técnicas é que SANS resulta de interações com o núcleo dos átomos, enquanto SAXS interage principalmente com a nuvem eletrônica (Svergun, 2007). Uma combinação entre essas técnicas pode, portanto, ser altamente complementar. No entanto, existem poucos reatores no mundo que permitem a utilização de SANS, sendo que não há nenhum no Brasil até o momento.

Figura 50. Semelhanças e diferenças entre as técnicas de cristalografia de raios-X e SAXS. Em A, uma representação simplificada das duas técnicas. Em B, um difratograma obtido por SAXS. Em C, um difratograma obtido por cristalografia. Observe que a figura ilustrada em B está aumentada com relação a C, sendo que, colocando-as na mesma escala, a figura B teria a dimensão do círculo vermelho indicado na figura C. Extraído da Página do Grupo coordenado por Nagar: <http://bhushan-nagar.lab.mcgill.ca/techniques.html>

4.1.3. Princípios da técnica de espalhamento de Raios-X a baixos ângulos (SAXS)

Raios-X são ondas eletromagnéticas nas quais a intensidade varia senoidalmente com o tempo. Esse campo magnético gerado provoca oscilações em partículas com carga, como elétrons. Por outro lado, a aceleração provocada no elétron faz com que ele emita uma onda eletromagnética secundária, esférica. No caso de um espalhamento elástico, o comprimento da onda emitida pelo elétron é igual ao comprimento de onda do feixe de raios-X incidente. Desta forma, como resultado disso, dizemos que elétrons “espalham” raios-X (Feigin e Svergun, 1987).

Figura 51. Esquema representando o espalhamento de raios x de uma partícula. Extraído de Feigin e Svergun (1987).

O problema do espalhamento de raios-X provocado por uma partícula, conforme ilustrado na figura 51, pode ser descrito através da equação de Schödinger, utilizando princípios de física quântica, sendo que uma onda $\Psi(\vec{r})$ espalhada por um campo $V(\vec{r})$ é dada pela equação 30 (Feigin e Svergun, 1987):

$$[\nabla^2 + k_0^2 + vV(\vec{r})]\Psi(\vec{r}) = 0 \quad \text{Equação 30}$$

Onde k_0 representa o número de onda da onda incidente, ∇^2 é o operador Laplaciano, e v é um parâmetro especificando a intensidade da interação com o campo potencial. Um dos

termos da solução da equação de Schödinger pode ser arranjada, gerando a equação 31 (Feigin e Svergun, 1987):

$$f(\vec{q}) = \frac{v}{4\pi} \int V(\vec{r}) \exp(i \vec{q} \cdot \vec{r}) d\vec{r} \quad \text{Equação 31}$$

O significado dos vetores está representado na figura 52. Vale notar que na figura, o ângulo de espalhamento é definido como θ , mas que muitas vezes ele é definido como 2θ .

Figura 52. Vetor da onda espalhada por uma partícula. Adaptado de Glatter e Krakty (1982).

A equação 32 é classificada como uma transformada de Fourier. A solução para o problema inverso é definida como (Feigin e Svergun, 1987):

$$V(\vec{r}) = \frac{1}{2\pi^2 v} \int f(\vec{q}) \exp(-i \vec{q} \cdot \vec{r}) d\vec{q} \quad \text{Equação 32}$$

Com isso, temos que a aplicação da transformada de Fourier nos permite calcular a amplitude de espalhamento de um sistema conhecido de centros espalhadores, e a transformada inversa nos permite, utilizando uma amplitude de espalhamento conhecida, calcular mais informações sobre os centros espalhadores.

4.1.3.1 Função dos pares de distância e suas aplicações para avaliar o formato de partículas

Na figura 53 está representado um esquema geral de como os dados de SAXS são coletados. Como pode ser observado, uma radiação monocromática de raios-X é incidida na amostra (ou no solvente) provocando um espalhamento, que é detectado em um detector bidimensional. Como amostras de proteínas globulares normalmente estão presentes em solução de forma randomicamente orientada, o resultado do espalhamento é isotrópico, ou seja, um círculo igual em todas as direções. Assim, é possível fazer a integração dos dados, de modo a obter um perfil de espalhamento em uma direção (Mertens e Svergun, 2010).

Figura 53. Esquema geral de um experimento de SAXS. Extraído de Mertens e Svergun (2010).

A equação 32 pode ser rearranjada de forma a obter-se a equação 33, que descreve a intensidade da onda espalhada (Feigin e Svergun, 1987; Oliveira, 2011).

$$I_1(q) = 4\pi \int_0^{D_{\max}} p(r) \frac{\sin qr}{qr} dr$$

Equação 33

Onde a $p(r)$ é chamada de função de pares de distância. Essa função descreve a probabilidade com que uma reta com comprimento r está contida dentro da partícula. A $p(r)$ é

uma função bastante informativa, pois permite avaliar prontamente o formato da partícula espalhadora. Esferas resultam em uma $p(r)$ em formato de sino, sendo que o valor de r correspondente ao máximo é o raio da esfera. Por outro lado, partículas alongadas tem um formato semelhante, mas o máximo é deslocado para a esquerda da curva. Partículas ocas, por outro lado, tem uma curva semelhante àquela de uma esfera, mas com máximo deslocado para a direita. Dímeros, por outro lado, apresentam uma $p(r)$ com duas regiões máximas (Figura 54).

Figura 54. Perfil de espalhamento esperado para partículas com diferentes formas geométricas. Em A, a intensidade de espalhamento. Em B, a $p(r)$. Extraído de Oliveira (2005) e (2011).

4.1.3.2 Lei de Guinier e sua importância para avaliar monodispersidade do sistema e ausência de interações entre partículas

Ao analisar espalhamentos de partículas, Guinier descreveu uma tendência linear no início da curva de intensidade de espalhamento (Guinier e Fournet, 1955). Essa tendência é obtida através de manipulação da equação 33. Essa equação pode ser expandida utilizando a expansão de McLaurin ($\sin(qr)/(qr) \approx 1 - (qr)^2/3! + \dots$), obtendo-se a equação 34.

$$I(q) = 4\pi \left[\int_0^{D_{\max}} p(r) dr - \frac{1}{6} q^2 \int_0^{D_{\max}} p(r) r^2 dr + \dots \right] \quad \text{Equação 34}$$

Que pode ser rearranjada na forma (equação 35):

$$I(q) = 4\pi \int_0^{D_{\max}} p(r) dr \left[1 - \frac{1}{3} \frac{\int_0^{D_{\max}} p(r) r^2 dr}{\int_0^{D_{\max}} p(r) dr} q^2 + \dots \right] \quad \text{Equação 35}$$

Sendo que o raio de giro pode ser definido pela equação 36, no contexto de SAXS:

$$R_g^2 = \frac{\int_0^{D_{\max}} p(r) r^2 dr}{2 \int_0^{D_{\max}} p(r) dr} \quad \text{Equação 36}$$

Com isso, substituindo-se a equação 36 na equação 35, é possível obter a equação 37, que pode ser simplificada para a equação 38 (Guinier e Fournet, 1955).

$$I(q) = I(0) \left[1 - \frac{1}{3} R_g^2 q^2 + O(q^4) + \dots \right] \cong I(0) \exp \left(-\frac{1}{3} R_g^2 q^2 \right) \quad \text{Equação 37}$$

$$\ln(I(q)) = \ln(I(0)) - \frac{1}{3} R_g^2 q^2 \quad \text{Equação 38}$$

Essa igualdade só é satisfeita para valores de q pequenos, via de regra, quando $R_g^2 q^2 < 1$. Isso porque, em valores menores do que 1, a quarta potência torna o número ainda menor, e com isso, os termos da expansão de quarta potência, ou maior, podem ser desprezados. No entanto, se q for um número grande, os termos de alta potência serão valores muito maiores, e consequentemente esses termos não poderão ser desprezados.

A Equação 38 é chamada de equação de Guinier, sendo que um plote de $\ln I(q)$ versus q^2 deve, segundo essa lei, resultar em uma reta para valores de q baixos (abaixo de 1, ou 1,3 como apontam Mertens e Svergun (2010)). A inclinação dessa reta será, portanto, $-1/3 R_g^2$, onde podemos, portanto, obter o valor de R_g . Caso esse plote não seja uma reta (Figura 55), isso é indicativo de o sistema não se encontra monodisperso, ou ainda que fatores de estrutura (*i.e.* interações entre partículas) podem estar comprometendo os dados (Mertens e Svergun, 2010).

Figura 55. Exemplo de plots de Guinier que desviam da linearidade. Extraído de Mertens e Svergun (2010).

4.1.3.3 Cálculo empírico da massa molecular da partícula

Além da análise de Guinier, a massa molecular da partícula pode ser calculada a partir do valor de $I(0)$. Para isso, os dados devem ser convertidos em escala absoluta. Isso pode ser feito de diversas maneiras, mas usualmente, para proteínas, é medido o espalhamento de água no capilar a ser utilizado, e a razão entre o espalhamento de água teórico (tabelado) e o espalhamento da água obtido é multiplicado pelos dados. Com isso, a equação 39 descreve como o valor de massa molecular pode ser obtido a partir dos dados de SAXS (Svergun *et al.*, 1998; Oliveira, 2005; Oliveira, 2011):

$$M_w = \frac{N_A \cdot I(0)}{c \Delta \rho_m^2} \quad \text{Equação 39}$$

Sendo que N_A é o número de Avogadro, $\Delta \rho_m$ é o contraste de espalhamento de massa da partícula e c a concentração. $\Delta \rho_m$ é, usualment, parecido em proteínas, e com isso podemos adotar seu valor de $2 \cdot 10^{10}$ cm/g para estimativa da massa molecular de proteínas no geral (Svergun *et al.*, 1998; Oliveira, 2011).

4.1.4 Estados oligoméricos de beta-glicosidases GH1

Apesar de muito conservadas em estrutura secundária e terciária, analisando a estrutura cristalográfica de 51 beta-glicosidases GH1, observa-se que 22 estruturas aparecem com uma única molécula de proteína na unidade assimétrica, sendo que as demais possuem 2, 3, 4, 6, 8 ou até 10 unidades por célula unitária (informação obtida através do site (Protein Data Bank)). Apesar da cristalografia nem sempre refletir o estado oligomérico da proteína em solução, como

discutido no item 4.1.2, este é um indício de que a estrutura quaternária de beta-glicosidases é bastante variada. Kim e Kim (2004), comparando diversas beta-glicosidases diméricas, observaram que a interface dimérica da linamarase de *Trifolium repens* é localizada na região C-terminal da enzima, enquanto que a região de interface da mirosinase de *Sinapis alba* é encontrada na região N-terminal, e a interface dimérica da beta-glicosidase Glu1 de *Zea Mays* é encontrada exatamente do meio da cadeia polipeptídica, o que revela uma grande diversidade na interface responsável pelos arranjos oligoméricos em beta-glicosidases GH1.

Além disso, utilizando técnicas que permitem avaliar a estrutura quaternária de beta-glicosidases em solução, alguns grupos têm encontrado diferentes formas oligoméricas para estas enzimas. Em um desses estudos, analisando a beta-glicosidase de *Saccharophagus degradans* por SAXS, Brognaro e colaboradores (2016) determinaram que em solução, a forma monomérica da enzima é predominante. Já em outro estudo, Colussi e colaboradores (2015), analisando duas beta-glicosidases de *Thermotoga petrophila*, observaram que ambas se apresentam como dímeros em solução. Por outro lado, em um trabalho que busca a correlação entre estados oligoméricos de beta-glicosidases e sua importância na função destas enzimas, Zanthorlin e colaboradores (2016) propuseram que o estado tetramérico da beta-glicosidase psicrófila de *Exiguobacterium antarcticum* B7 é importante para a atividade em baixas temperaturas, possivelmente por uma estabilização da estrutura nativa a baixas temperaturas promovida por este arranjo oligomérico. Desta forma, o arranjo tetramérico desta beta-glicosidase parece estar envolvido com a adaptação ao frio, juntamente com uma redução do número de pontes salinas, um aumento na exposição de grupos hidrofóbicos, e um aumento da flexibilização de regiões expostas ao solvente (Zanthorlin *et al.*, 2016).

Além disso, Kim e colaboradores (2000) observaram um arranjo oligomérico não muito usual ao analisar a beta-glicosidase de aveia, a qual forma arranjos fibrilares, sendo possível observar por microscópio eletrônico de transmissão que as fibrilas se apresentam em diferentes tamanhos. Essas fibrilas são organizadas por um arranjo de monômeros de 60 kDa formando um anel com, possivelmente, 6 subunidades, que se empilham formando fibrilas com milhões de kDa.

Enfim, observa-se uma grande variedade de estados oligoméricos em beta-glicosidases, mesmo que estas enzimas sejam bastante conservadas em estrutura terciária. Apesar de alguns trabalhos sugerirem que esse arranjo pode ter relação com a função destas proteínas (Zanphorlin *et al.*, 2016), ainda há muito a se investigar sobre a relação entre a estrutura oligomérica de beta-glicosidases e sua importância para as propriedades enzimáticas.

4.1.5 Uso de redes para avaliar estrutura quaternária de proteínas

A análise de Redes de Estrutura Proteica tem se revelado um método interessante para a elucidação de resíduos constituintes de rotas alostéricas (Del_Sol *et al.*, 2006). No entanto, pouco se sabe sobre os possíveis papéis dos resíduos centrais da Rede de Estrutura Proteica na manutenção do estado oligomérico de proteínas.

Em 2005, del Sol e O'Meara analisaram diversos complexos proteicos como redes, onde os resíduos são os nós e a interação entre eles conexões, que podem ser intra ou inter-cadeias polipeptídicas. Os autores observaram que os resíduos centrais por *betweenness* do complexo frequentemente correspondem ou estão em contato com *hotspots* de interação entre as cadeias polipeptídicas, indicando que esses resíduos centrais podem desempenhar papéis importantes para a formação do arranjo oligomérico de proteínas (Del Sol e O'meara, 2005).

Contudo, poucos trabalhos têm conseguido utilizar a abordagem em redes para caracterizar a propagação de informação ao longo de estruturas proteicas oligoméricas.

4.2 Materiais e Métodos

4.2.1 Análise por Espalhamento de Luz dinâmico (DLS).

Inicialmente, amostras de Sfβgli produzidas em *E. coli* BL21 gold (DE3), purificadas, homogêneas de acordo com análise em SDS-PAGE (capítulo 1), tiveram a sua monodispersidade avaliada por DLS. Para isso, 5 µL de amostra foram colocadas dentro de cubetas acrílicas e então, o espalhamento de luz dinâmico das amostras foi medido no equipamento 90 Plus da *Brookhaven Instruments*, instalado no Instituto de Física da USP. Após a aquisição, a curva de correlação foi ajustada utilizando-se o modelo CONTIN e a partir disso as distribuições em intensidade, volume e número com relação ao diâmetro de partículas foram produzidas.

4.2.2 Coleta de dados de espalhamento de raios-X a baixos ângulos

Para avaliar o perfil oligomérico da Sfβgli, amostras contendo a enzima purificada foram concentradas utilizando diálise reversa com polietilenoglicol (PEG) 35000 até atingir a concentração necessária para os experimentos de SAXS. Após isso, o espalhamento de raios-X das amostras foi medido com o apoio do Prof. Dr. Cristiano Oliveira Pinto, Dra. Renata Bicev e Dennys Reis (Instituto de Física – USP). Para isso foram utilizados os equipamentos *NanoStar* e *XEUSS*, ambos localizados no Instituto de Física – USP. Os dados de espalhamento foram acumulados por 30 min, sendo que para amostras com concentrações menores do que 2,0 g·L⁻¹, foram feitas 4 medições. A curva final analisada foi a média entre as 4 medições. Já em

concentrações mais altas, duas medições foram feitas, sendo a curva de espalhamento final a média entre as duas medições.

Além das amostras, foram determinados os espalhamentos de água e do tampão utilizado na análise. Para normalização dos dados para escala absoluta, uma razão entre o espalhamento da água teórico (tabelado, em cm^{-1}) e experimental foi calculado e multiplicado por todos os dados.

4.2.3 Transformada inversa de Fourier

Para produzir a transformada inversa de Fourier, os dados normalizados foram tratados utilizando o programa *GNOM*, disponível gratuitamente no pacote *ATSAS* (<https://www.embl-hamburg.de/biosaxs/software.html>). Para o cálculo da função de pares $p(r)$, foram consideradas as opções padrão do programa, sendo que o sistema foi considerado monodisperso, e o espalhamento foi considerado como sem efeito de *smearing*. O diâmetro máximo da partícula foi ajustado para uma melhor modelagem da curva de $p(r)$. Com isso, um arquivo de extensão “.log” foi gerado contendo as funções de pares das amostras testadas. Somente as amostras em fosfato de potássio 10 mM pH 6,0 (F10) não foram analisadas através do programa *GNOM*, pois o ajuste não foi possível, provavelmente devido a não monodispersidade do sistema ou ao grande tamanho das partículas em solução.

4.2.4 Análise gráfica dos dados

Para avaliar o sistema, plotes de Guinier foram produzidos com os dados de espalhamento. Para isso, gráficos do logaritmo natural da intensidade espalhada [$\ln I(q)$] versus o

ângulo de espalhamento ao quadrado [q^2] foram produzidos, e a sua linearidade em baixos ângulos foi verificada. Além disso, o raio de giro (R_g) foi calculado a partir da inclinação da reta, segundo a Equação 38. A intensidade a $q \rightarrow 0$ também foi determinada pela reta, através do intercepto do ajuste linear de Guinier em $q=0$. Com este dado, o cálculo da massa molecular da partícula foi feito através da Equação 39.

Além disso, plotes de Krakty foram produzidos para avaliar a flexibilidade do sistema. Para isso, gráficos de Iq^2 versus q foram produzidos. Neste gráfico, a presença de uma curva de sino no início dos plotes indica a presença de uma proteína globular nas amostras. Por outro lado, um aumento de sinal para altos valores de q indica flexibilidade do sistema.

4.2.5 Produção de um modelo utilizando o programa DAMMIF

Para produção de um modelo do envelope da partícula, foi empregado o programa DAMMIF. Para isso, o programa foi acessado gratuitamente na plataforma *online* do ATSAS (<https://www.embl-hamburg.de/biosaxs/atsas-online/dammif.php>). A curva gerada pelo programa GNOM, segundo o item 4.2.3, foi utilizada no programa DAMMIF, considerando que não há simetria e que a anisotropia é desconhecida. O programa foi executado por 20 vezes, onde, por fim, através do programa DAMAVER, um modelo de envelope médio foi gerado. Para sobrepor este modelo à estrutura em pdb, o programa *suppdb*, do pacote ATSAS, foi utilizado.

4.2.6 Produção de curvas de espalhamento esperadas para estruturas em pdb

Para gerar curvas de espalhamento teóricas a partir de estruturas cristalográficas depositadas no PDB, foi empregado o programa Crysol. Para isso, o programa, que é disponível

no pacote ATSAS (<https://www.embl-hamburg.de/biosaxs/software.html>), foi executado no modo de análise de amplitudes de espalhamento e envelope. A partir disso, a estrutura em formato pdb foi utilizada no programa como ponto de partida para a geração de dados de espalhamento. Os parâmetros usados foram os padrões do programa (15 *maximum order of harmonics*, 17 *order of Fibonacci grid*, valor de q máximo de 0,5, 51 pontos na curva gerada e não considerar hidrogênios explícitos). Com base nisso, curvas de intensidade de espalhamento em função de q foram geradas, e comparadas com valores experimentalmente obtidos. No caso do espalhamento esperado para a cadeia monomérica de Sf β gli, como não há uma estrutura resolvida onde esta enzima apareça como monômero, considerou-se somente a cadeia A da estrutura cristalográfica 5CG0 como partícula espalhadora. Já para gerar o perfil de espalhamento esperado para o dímero, as cadeias A e B desta estrutura cristalográfica foram utilizadas. Inclusive, para construir um dímero com distâncias entre as cadeias diferentes daquela observada na estrutura cristalográfica 5CG0 foi utilizado o programa DIMFOM, disponível no pacote ATSAS. Similarmente, no caso do trímero, tetrâmero e hexâmero foram empregadas as cadeias B-E-F, A-B-E-F, e A-B-C-D-E-F, respectivamente. Para produção do perfil de espalhamento do dodecâmero, um modelo de dodecâmero de Sf β gli foi previamente construído utilizando o programa SASRef (item 4.2.7), o qual foi considerado como partícula espalhadora.

4.2.7 Modelagem de corpos rígidos utilizando o programa SASRef.

Para produção de um modelo do dodecâmero de Sf β gli, uma modelagem de corpos rígidos foi feita utilizando o programa SASRef. Para isso, o programa, que é disponível no pacote

ATSAS (<https://www.embl-hamburg.de/biosaxs/software.html>) foi usado no modo usuário. O ajuste foi feito considerando-se 2 subunidades, sendo que cada uma delas correspondia à estrutura cristalográfica 5CG0, contendo 6 cadeias polipeptídicas. Somente simetria P1 foi considerada nesse processo. Por fim, o ajuste foi feito buscando encontrar um modelo que melhor se ajustasse aos dados da curva experimental de espalhamento gerada pela amostra de S β gli em tampão fosfato de potássio 10 mmol·L⁻¹ pH 6,0 (F10).

4.2.8 Análise dos dados utilizando o programa *Ensemble Optimization Method* (EOM)

Para avaliação mais detalhada da flexibilidade dos diferentes oligômeros, as curvas experimentais de espalhamento foram analisadas pelo programa EOM, através do servidor online: <https://www.embl-hamburg.de/biosaxs/atsas-online/eom.php>. Apenas simetria do tipo P1 foi considerada. Para produção do arquivo de sequência, foi utilizado o programa pdb2seq, disponível no pacote ATSAS (<https://www.embl-hamburg.de/biosaxs/software.html>). Para isso na *prompt* de comando do Windows, gerada através de um atalho *cmd* na própria pasta do pacote ATSAS, foi executado o comando: *Pdb2seq nomedoarquivo.pdb>nomedoarquivo.seq*

Assim, na própria pasta do pacote ATSAS, é gerado um arquivo com extensão .seq, que pode ser utilizado no programa EOM.

Para avaliação da flexibilidade relativa das unidades do dímero, um arquivo pdb contendo as cadeias A e B da estrutura cristalográfica 5CG0 foi utilizado como ponto de partida para a produção do arquivo “.seq”. No entanto, para possibilitar flexibilidade relativa entre as unidades, um *linker* contendo repetições de treonina e prolina foi inserido entre as sequências

das duas cadeias polipeptídicas. Por fim, foram considerados modelos com estruturas do tipo nativa, assumindo-se a presença de dois domínios. A estrutura de cada um desses domínios foi considerada como sendo aquela da cadeia A da Sf β gli obtida por cristalografia.

4.2.9 Análise dos dados utilizando o programa *Oligomer*

Para calcular a abundância relativa de cada tipo de oligômero de Sf β gli em solução, foi utilizado o programa *Oligomer*, disponível no pacote ATSAS (<https://www.embl-hamburg.de/biosaxs/software.html>). Assim, inicialmente empregando-se o programa *FFmaker*, também disponível no mesmo pacote, foram reunidos em um único arquivo os perfis de espalhamento esperados para as diferentes formas oligoméricas gerados com o programa *Crysol* (item 4.2.6). Com base nisso, o programa *Oligomer* foi executado no modo 0 (conjunto de fatores de forma e diversos conjuntos de dados experimentais). Como fator de forma foi utilizado o arquivo de extensão .dat gerado pelo programa *FFmaker*. Para melhor ajuste dos dados no programa *Oligomer*, a constante padrão apresentada pelo programa foi subtraída dos dados. Por fim, um ajuste dos dados com os perfis de espalhamento de cada uma das formas oligoméricas foi apresentado pelo programa *Oligomer*, o qual continha as frações volumétricas (abundâncias relativas) de cada forma oligomérica de Sf β gli presente em solução.

4.2.10 Ensaio de cinética enzimática de Sf β gli

Os ensaios de atividade da beta-glicosidase Sf β gli foram realizados utilizando 50 μ L de amostra contendo a enzima diluída 20x (em CF100 ou em F10) e 50 μ L do substrato *p*-nitrofenil- β -D-glucopiranosídeo (*p*NPGLu) em diversas concentrações preparado em tampão citrato-fosfato

100 mmol·L⁻¹ pH 6,0 (CF100). As amostras de Sfβgli foram previamente preparadas em tampão CF100 ou F10. A mistura de reação foi incubada em diferentes temperaturas (de 28°C a 40°C) por intervalos de tempo de 5, 10, 15 e 20 minutos e em seguida interrompida com 100 μL de solução de carbonato de sódio 0,5 mol·L⁻¹. Com isso, o *p*-nitrofenolato formado foi detectado por meio da absorbância em 415 nm usando um leitor de placas Elx800 (Biotek®). As velocidades iniciais destas reações foram calculadas e ajustadas na equação de Michaelis-Menten (Equação 20), de onde foram obtidos os parâmetros de V_{max} e K_m .

4.2.11 Determinação da centralidade dos resíduos de diferentes estados oligoméricos da Sfβgli

Para produzir as redes de estrutura proteica o arquivo em formato PDB da Sfβgli (5CG0) foi aberto utilizando o programa *CMView1.1.1*. Para gerar a rede dos oligômeros, os resíduos das cadeias da Sfβgli foram renumerados em modo contínuo criando um arquivo pdb contendo somente uma cadeia que englobava todas as cadeias da estrutura cristalográfica (total = 2036 resíduos). Para viabilizar o processamento no programa *CMView1.1.1*, foi empregado como critério de conexão apenas a distância entre os carbonos-alfa, estabelecendo como corte a distância de 8Å. Após a produção da rede estrutural pelo programa, um mapa de contatos foi exportado. A partir deste, a rede foi caracterizada utilizando o programa *Cytoscape 3.2.0*, analisando-a utilizando o comando *Analyze Network*. Uma planilha foi gerada contendo a centralidade de cada um dos resíduos por *closeness*, *betweenness* e grau, bem como os parâmetros gerais da rede.

4.3 Resultados e Discussões

4.3.1 Determinação do estado oligomérico da Sf β gli em solução

A beta-glicosidase de *Spodoptera frugiperda* (Sf β gli) foi cristalizada recentemente, revelando 6 unidades monoméricas por célula unitária (Tamaki *et al.*, 2016), conforme a figura 56.

Figura 56. Estrutura cristalográfica da Sf β gli (5CG0). Extraído de Tamaki *et al.* (2016).

No entanto, assim como já relatado para outras beta-glicosidases, este arranjo pode não ser igual ao encontrado em solução. Neste sentido, foram feitos experimentos de espalhamento de raios-X a baixos ângulos (SAXS) da amostra em solução, para avaliar qual seria o arranjo oligomérico preponderante nesta condição. Para isso, amostras de Sf β gli em tampão citrato-fosfato 100 mM pH 6,0 (CF100) e em tampão fosfato de potássio 10 mM pH 6,0 (F10) foram inicialmente analisadas por DLS (*Dynamic Light Scattering*) para avaliar a sua monodispersidade. Esta pré-avaliação permite determinar se as amostras estão adequadas para a análise por SAXS. A coleta em dois tampões diferentes permite ainda avaliar se existem mudanças no perfil

oligomérico da enzima de acordo com o tampão ou força iônica. Na figura 57 estão os resultados obtidos para as amostras em CF100 e F10 indicando que na análise por intensidade de sinal há mais de um pico sugerindo a presença de mais de uma população de partículas com tamanhos distintos na amostra. No entanto, como o sinal cresce com o quadrado do volume, ao normalizar esses dados em volume e em número de partículas, observa-se que quase todas as partículas da amostra possuem o mesmo tamanho, indicando que estão monodispersas em solução, o que é uma informação prévia importante para a análise por SAXS. Além disso, é curioso notar que em F10 o diâmetro das partículas parece ser cerca de 6x maior do que em CF100 (aproximadamente 400 Å e 70Å, respectivamente). Isso pode indicar que o estado oligomérico da enzima nestes tampões é diferente, ou ainda que fatores de estrutura, como atração/repulsão entre partículas, podem estar afetando os dados, gerando um falso diâmetro de partícula em um dos tampões utilizados.

Com base nestas observações, para avaliar o estado oligomérico da Sfβgli nos dois tampões, experimentos de SAXS foram realizados com apoio do Prof. Dr. Cristiano Oliveira Pinto e seu grupo no IF USP. Na figura 58 estão os resultados de espalhamento obtidos para a Sfβgli em tampão CF100 e F10. Observa-se que, apesar da concentração das amostras ser semelhante, há uma nítida diferença no padrão de espalhamento.

A partir destes resultados, a tendência linear de Guinier foi analisada, para verificar a presença de efeitos de atração/repulsão entre as partículas. Na figura 59 observa-se que o plote de Guinier para Sfβgli em tampão CF100 apresenta uma tendência linear para baixos valores de q , o que indica que a amostra está monodispersa e sem fatores de atração/repulsão significativos entre as partículas.

Figura 57. Perfis de espalhamento dinâmico de luz (DLS) da Sfβgli selvagem em tampão CF100 e F10. Em A e B estão as curvas de correlação obtidas por DLS. Em C e D estão os perfis de distribuição derivados dos dados de DLS. Em A e C, resultados obtidos em tampão CF100. Em B e D, em tampão F10.

Por outro lado, o plote de Guinier para Sfβgli em tampão F10 não apresenta a mesma tendência, exibindo um desvio de linearidade caracterizado por intensidades mais altas do que o esperado em baixos valores de q . Isso é indicativo de efeitos de atração entre as partículas, possivelmente pela baixa força iônica do tampão, ou ainda da presença de agregados de maior massa molecular na amostra. Utilizando a inclinação desta reta, é possível estimar o raio de giro médio das partículas (R_g). No caso, temos que a Sfβgli em tampão CF100 possui R_g de aproximadamente 40 Å, enquanto que a enzima em tampão F10 possui R_g de 117 Å. Isso indica

que, a amostra em F10 é composta por partículas bem maiores do que em CF100, de forma coerente com os resultados de DLS.

Figura 58. Perfil de espalhamento por SAXS da SfbgI selvagem, a 25°C e pH 6,0. Em vermelho, dados coletados em tampão fosfato de potássio 10 mM (F10); em preto, dados coletados em tampão citrato-fosfato 100 mM (CF100). A concentração das duas amostras utilizadas para a análise foi de 0,8 mg/mL.

Figura 59. Plote de Guinier obtido para a SfbgI selvagem em tampão CF100 (preto) e F20 (vermelho).

Para melhor caracterizar estas diferenças de R_g segundo à condição do meio, utilizando a estrutura cristalográfica da SfbgI juntamente com o programa Crysol, foram determinados os R_g

prováveis para as estruturas monomérica (cadeia A), dimérica (cadeias A e B), trimérica (cadeias B, E e F), tetramérica (cadeias A, B, E e F) e hexamérica (cadeias A, B, C, D, E e F). Os valores obtidos foram 21,6 Å, 32,8 Å, 42,7 Å, 54,8 Å e 78,0 Å, respectivamente. Comparando estes dados com os obtidos por SAXS, temos que a amostra em CF100 tem R_g entre um dímero e um trímero de Sfβgli, e em F10 o R_g é maior do que da estrutura hexamérica.

Com base nisso, utilizando a intensidade de raios-X espalhados a ângulos tendendo a 0, $I(q=0)$, e sabendo que a concentração das duas amostras foi de 0,8 mg/mL, foi calculada a massa molecular das partículas presentes nos tampões CF100 e F10 utilizando a Equação 39. Desta forma, temos que a partícula na amostra em CF100 possui um valor de 135 kDa, enquanto que na amostra em F10 possui um valor estimado de 401 kDa. Como a Sfβgli monomérica possui massa molecular de 58 kDa, temos que em CF100 o peso parece corresponder a um dímero da Sfβgli, enquanto que em F10, o peso corresponderia a um hexâmero ou até octâmero.

Desta forma, foi possível observar uma clara diferença no estado oligomérico da Sfβgli nos dois tampões testados, sendo que em CF100, que possui maior força iônica, a proteína parece estar na forma dimérica e em F10 estaria em estados oligoméricos maiores que poderiam ser até mesmo octâmeros.

4.3.2 Análise por SAXS da Sfβgli em tampão CF100

Para uma análise mais acurada do estado oligomérico da Sfβgli em tampão CF100, amostras com diferentes concentrações de enzima foram obtidas. Então, os perfis de espalhamento das amostras foram analisados plotando-se gráficos de Guinier (Figura 60) (Guinier e Fournet, 1955). Pode ser observado que em concentrações de Sfβgli acima de 2,1

mg/mL ocorre um desvio de linearidade na região de Guinier, indicando que as partículas estão em processo de atração, formando agregados. No entanto, considerando a região de linearidade do plote de Guinier destas amostras, é possível estimar o $I(q=0)$ desconsiderando a presença desses agregados. Na figura 61 observa-se que este valor aumenta linearmente com a concentração de proteína. Isso indica que a concentração não altera de forma significativa o estado oligomérico da proteína.

Para comparar a curva de espalhamento obtida com a curva esperada caso a proteína fosse um monômero, dímero, trímero, tetrâmero e hexâmero, a estrutura cristalográfica da Sfβgli (5CG0) foi utilizada como objeto espalhador no programa CRY SOL. Este programa gera a curva de espalhamento de SAXS esperada para a estrutura proteica em questão. Desta forma, para simular diferentes estados oligoméricos (dímero, trímero, tetrâmero e hexâmero), foram feitos arranjos entre as 6 cadeias que fazem parte da estrutura cristalográfica da Sfβgli conforme já comentado acima. Para simular o monômero, foi considerada somente a cadeia A.

Na figura 62 estão representadas as curvas de espalhamento esperadas para monômero, dímero, trímero, tetrâmero e hexâmero. Além disso, os dados experimentais coletados com amostra de Sfβgli em concentração de 1,6 mg/mL estão também representados. Observa-se uma grande sobreposição dos dados experimentais com o resultado esperado para um dímero, indicando que a proteína em solução está presente na forma dimérica. Esta observação é coerente com a massa molecular calculada acima (item 4.3.1) para Sfβgli em CF100.

Ajuste linear – Plote de Guinier

Para concentração 0,8 mg/mL: $\ln(I) = (-450 \pm 40) \times q^2 + (-2,76 \pm 0,03)$

1,6 mg/mL: $\ln(I) = (-440 \pm 30) \times q^2 + (-2,48 \pm 0,03)$

2,1 mg/mL: $\ln(I) = (-380 \pm 20) \times q^2 + (-2,14 \pm 0,02)$

5,2 mg/mL: $\ln(I) = (-363 \pm 13) \times q^2 + (-1,52 \pm 0,01)$

Figura 60. Curva de espalhamento de raios-X a baixos ângulos. (A) curva de intensidade de espalhamento em função de q . (B) Plote de Guinier. Em ambos os gráficos, os pontos cinza foram obtidos com uma concentração de Sfβgli de 0,8 mg/mL, os pontos vermelhos com concentração de 1,6 mg/mL, os pontos azuis com 2,1 mg/mL e os pontos verdes, com concentração de 5,2 mg/mL.

Figura 61. Relação entre a intensidade quando o ângulo tende a 0 - $I(0)$ – no plote de Guinier e a concentração de Sf β gli. A linha em preto é a regressão linear dos dados.

Figura 62. Comparação entre o espalhamento experimentalmente observado para a Sf β gli em CF100 e os perfis esperados para diferentes formas oligoméricas da enzima. Em A, a curva em escala linear e em B, a curva em escala logarítmica. Sf β gli em concentração de 1,6 mg/mL.

Com base nesses dados, a transformada de Fourier inversa foi calculada utilizando-se o programa GNOM. Através das funções de pares $p(r)$ ilustradas na figura 63, observa-se que não há mudanças no R_g da amostra em diferentes concentrações, reforçando a hipótese de que não

há mudanças significativas no estado oligomérico nas concentrações analisadas por SAXS. Interessantemente nota-se a presença de ombros nas curvas, o que sugere que a amostra possui algum tipo de oligomerização, possivelmente dimérica com base no R_g .

Figura 63. Curva da função dos pares de distância $[p(r)]$ das amostras de Sfβgli em diferentes concentrações, todas em tampão CF100. Dados obtidos através do programa GNOM.

Buscando análises adicionais para identificar o estado oligomérico que melhor se ajustava aos dados experimentais, as curvas esperadas para cada tipo de estado oligomérico da Sfβgli obtidas pelo programa CRY SOL (Figura 62) foram utilizadas para gerar a Transformada de Fourier inversa, obtendo-se a $p(r)$ de cada estado oligomérico através do programa GNOM. Na figura 64 estão ilustradas as curvas normalizadas, juntamente com a curva de $p(r)$ obtida com os dados experimentais da Sfβgli a $1,6 \text{ mg/mL}$. Mais uma vez, observa-se grande sobreposição dos dados experimentais aos esperados para o dímero da Sfβgli.

Figura 64. Função dos pares $[p(r)]$ de diversos estados oligoméricos da Sfβgli.

Dessa forma frente às evidências da presença de um dímero de Sfβgli em solução, para determinar o seu formato e confrontá-lo com o dímero formado pelas cadeias A e B da estrutura cristalográfica de Sfβgli, o perfil de espalhamento da amostra com concentração 1,6 mg/mL foi utilizado para produção de um modelo *ab initio* de baixa resolução com o software DAMMIF. Buscando um melhor ajuste dos dados, foram produzidos diversos modelos, e em seguida um modelo médio foi obtido utilizando o programa DAMAVER. Por fim, este modelo final foi sobreposto com a estrutura do dímero obtido pela cadeia A e B da estrutura cristalográfica (5CG0; Tamaki *et al*, 2016). Como pode ser observado na figura 65, as dimensões do modelo *ab initio* são parecidas com as dimensões do dímero cristalográfico. No entanto, na estrutura cristalográfica, as duas unidades do dímero parecem um pouco mais próximas do que o observado no modelo *ab initio*.

Figura 65. Duas perspectivas do modelo *ab initio* gerado pelo programa DAMMIF em sobreposição com a estrutura cristalográfica de Sfβgli. As cadeias A e B da estrutura cristalográfica de Sfβgli estão representadas em *cartoon*, enquanto o modelo *ab initio* está representado em superfície cinza.

Para avaliar a distância das unidades que seria melhor ajustada, um novo modelo da estrutura do dímero foi criado utilizando o programa DIMFOM. Para isso, 2 unidades monoméricas de Sfβgli foram utilizadas, ambas a cadeia A, e o programa encontrou a posição relativa entre elas que melhor se ajustaria nos dados experimentais de espalhamento de SAXS. O resultado obtido está ilustrado na figura 66, sendo que a sua estrutura foi sobreposta ao modelo *ab initio* obtido pelo DAMMIF. Observa-se uma melhor sobreposição entre os modelos, sugerindo que as unidades de Sfβgli formadoras do dímero em solução podem de fato estar mais afastadas do que na estrutura cristalográfica.

Figura 66. Modelo de um dímero da Sfβgli. O dímero representado em *cartoon* foi obtido através do software DIMFOM utilizando por duas vezes a cadeia A do modelo cristalográfico da Sfβgli. A superfície em cinza representa o modelo obtido pelo programa DAMMIF.

No entanto, também não deve ser descartado que poderia haver flexibilidade na distância de separação entre as unidades formadoras do dímero, o que geraria distorção no modelo *ab initio*, pois a curva de SAXS é a somatória do espalhamento entre todas as conformações existentes. De fato, ao comparar os dados experimentais de espalhamento com o perfil esperado para o dímero da cadeia A e B da estrutura cristalográfica ou para o dímero resultante do DIMFOM (Figura 67), observa-se que todos ficam muito próximos entre si, e que os dados experimentais ajustam-se melhor a um modelo em determinados trechos da curva e a outro em outros trechos. Com isso, é possível que os dímeros em solução realmente apresentem flexibilidade na distância de separação entre as cadeias, o que faz com que o espalhamento de SAXS seja uma média das diferentes conformações.

Figura 67. Perfil de espalhamento de modelos de dímero de Sfβgli obtidos a partir da estrutura cristalográfica (curva preta) e pelo programa DIMFOM (curva vermelha). Os dados experimentais estão representados por pontos brancos.

Assim, para analisar a flexibilidade da Sfβgli em solução, foram feitos plotes de Krakty (Oliveira, 2005; Mertens e Svergun, 2010) de cada uma das curvas de espalhamento obtidas para amostras em CF100. Observa-se na figura 68 que a curva de Krakty possui um trecho em forma de curva de sino, indicando que a proteína em solução apresenta-se na forma globular, conforme o esperado. No entanto, após este trecho, nota-se uma elevação significativa na curva, sendo que $I \cdot q^2$ assume valores bastante altos em q maiores. Isso sugere a presença de trechos flexíveis na proteína. Adicionalmente observa-se que o plot de Krakty esperado para o dímero obtido da estrutura cristalográfica, pensando em um modelo estático sem nenhum tipo de flexibilidade, não apresenta esse aumento linear substancial de $I \cdot q^2$ para altos valores de q . Apesar de não haver um critério claro para definição de flexibilidade pelo plot de Krakty, esta análise indica que o dímero tem uma maior flexibilidade em solução do que no modelo

cristalográfico, sendo que essa flexibilidade poderia ocorrer justamente na junção da estrutura dimérica levando a uma flutuação na distância entre os monômeros.

Figura 68. Comparação entre o plote de Kratky do modelo dimérico da SfbgI com o plote produzido a partir dos dados experimentais.

Desta forma, para estimar a flexibilidade do sistema, além da análise de Kratky, o espalhamento da amostra da SfbgI também foi analisado pelo programa *Ensemble Optimization*

Method 2.1 (EOM) (Tria *et al.*, 2015), que cria diversas posições relativas entre as unidades constituintes do dímero, e seleciona as posições mais prováveis com base no perfil de espalhamento experimental. Através deste programa, observa-se que a população de estruturas diméricas mais provável em solução possui uma distribuição mais discreta do que a distribuição de todas as conformações possíveis (Figura 69).

Além disso, nota-se que há duas populações de moléculas com R_g bastante distintos entre si. Integrando-se os picos, obtemos uma estimativa de que aproximadamente 88% das partículas encontram-se em uma conformação com R_g 34,6 Å, enquanto que 12% das moléculas encontram-se em conformações com R_g de aproximadamente 60 Å. É curioso notar que o segundo grupo de partículas têm raio de giro cerca de 2x maior do que o primeiro grupo. De fato, o R_g da primeira população é muito próximo ao R_g esperado para o dímero observado na estrutura cristalográfica de Sfβgli (32,8 Å), enquanto que a segunda população tem R_g próximo ao do tetrâmero observado na cristalografia (54,8 Å). A variação em torno de 34,6 Å observada no R_g do dímero (Figura 69), parece ser compatível com a flexibilidade e flutuação na distância entre os monômeros formadores do dímero discutida acima no caso do plote de Krakty e dos modelos gerados pelo DAMMIF e DIMFOM.

Ao observar os modelos estruturais mais prováveis gerados pelo programa EOM, observa-se que dos 4 modelos gerados, 3 são muito próximos entre si (Figura 69B, estruturas amarela, verde e azul), com distância entre cadeias similar à distância do dímero observado na estrutura cristalográfica de Sfβgli. Por outro lado, um dos modelos gerados difere bastante dos demais por apresentar as duas cadeias bastante distantes entre si (Figura 69B, estrutura rosa), o qual pode representar o tetrâmero.

Figura 69. Modelos de dímeros de Sfβgli gerados pelo programa EOM. Em A, a distribuição de R_g dos 10000 modelos inicialmente gerados (azul) e dos modelos selecionados (preto). Em B, a estrutura dos 4 modelos mais prováveis selecionados pelo programa, cada um em uma cor diferente (verde, amarelo, azul e rosa).

Como o programa do EOM não descreve o equilíbrio entre as diferentes formas oligoméricas, os dados de espalhamento foram utilizados no programa OLIGOMER para avaliar a quantidade relativa de cada estado oligomérico de Sfβgli em solução. Desta forma, assumindo

que no sistema haveria somente dímeros e tetrâmeros, o programa OLIGOMER estimou uma fração volumétrica de 91 ± 1 % de dímeros e 9 ± 1 % de tetrâmeros. Esses valores são muito próximos à distribuição dos R_g obtidos com o programa EOM. Portanto, este conjunto de resultados indica a presença de dímeros e tetrâmeros de SfbgIi em solução, sendo que a concentração de tetrâmeros é bem mais baixa do que a de dímeros.

Uma análise, utilizando o programa OLIGOMER, com as curvas de espalhamento geradas em diferentes concentrações de SfbgIi indica ainda que a abundância relativa de dímeros e tetrâmeros não varia muito com a concentração de proteína na faixa de 0.8 - 5 mg/mL (Figura 70), o que é coerente com os resultados inicialmente discutidos, de que a concentração não afeta significativamente o espalhamento (Figura 61).

Figura 70. Fração volumétrica de dímeros e tetrâmeros de SfbgIi em tampão CF100 estimada pelo programa OLIGOMER.

Por fim, vale destacar que as amostras utilizadas em SAXS possuem frequentemente concentrações de proteína maior do que o encontrado em sistemas *in vivo*. Neste sentido, a

presença de dímeros e tetrâmeros de Sfβgli nos experimentos de SAXS pode não refletir as condições biológicas. No entanto, Marana e colaboradores (2000) observaram anteriormente, que a Sfβgli extraída diretamente das larvas da lagarta *Spodoptera frugiperda* possuía peso molecular relativo compatível com um dímero em análises de cromatografia por gel filtração. Isso sugere que esta forma dimérica pode estar presente *in vivo*.

4.3.3 Análise por SAXS da Sfβgli em tampão F10

Conforme relatado no item 4.3.1, a Sfβgli em tampão F10 apresentou um perfil de espalhamento de SAXS significativamente diferente da amostra em CF100 (Figura 58). Além disso, o plote de Guinier sugeriu a presença de grandes agregados (Figura 59).

Analizando o perfil de espalhamento no programa GNOM, não foi possível determinar a curva $p(r)$ dos dados experimentais, possivelmente pela polidispersidade do sistema ou pelo tamanho acentuadamente grande das partículas, o que acarreta em ausência de um número significativo de dados com baixos valores de q . No entanto, comparando o perfil de espalhamento da amostra em F10 com os espalhamentos esperados para monômero, dímero, trímero, tetrâmero, hexâmetro e dodecâmero da Sfβgli simulados pelo programa CRY SOL, nota-se que o espalhamento experimental é mais próximo ao de um dodecâmero (Figura 71). Isso indica que a Sfβgli em solução com baixa força iônica possui uma forma oligomérica bastante grande.

Figura 71. Comparação entre o perfil de espalhamento calculado pelo CRY SOL para diferentes formas oligoméricas de Sfβgli e o espalhamento experimental obtido com a enzima 0,8 mg/mL em F10.

Desta forma, como o resultado de espalhamento sugere a presença de grandes oligômeros, o aumento de intensidade a baixos valores de q , que ocorre no plote de Guinier (Figura 59), pode ser compatível com estes grandes oligômeros presentes na solução. De fato, olhando o trecho inicial do plote de Guinier, observa-se que há presença de uma reta em baixos valores de q . Extrapolando o valor desta reta para $q=0$, pode-se estimar a massa molecular correspondente aos oligômeros que geram estes dados. Este valor foi de 401 kDa, que seria compatível com 7 a 8 unidades de Sfβgli constituindo o oligômero. Já, o valor de R_g estimado foi de 117 Å, um valor bastante elevado. Contudo estes parâmetros devem ser tomados com cuidado visto que para a linearidade de Guinier ser satisfeita, $R_g \times q$ deve ser menor ou igual 1,3 (Mertens e Svergun, 2010). Desta forma, neste caso somente valores muito pequenos de q

podem ser utilizados na composição da linearidade de Guinier, o que faz com que somente os 3 primeiros pontos da curva possam ser utilizados nessa análise. Com um número de dados bastante reduzido pode ocorrer grande erro na estimativa do $I(0)$, e portanto na massa molecular, e também no R_g , que está relacionado a inclinação desta reta.

De qualquer forma, observa-se que em F10, um tampão de baixa força iônica, há presença de oligômeros combinando várias unidades de Sfβgli.

4.3.4 Efeito do estado oligomérico na atividade de Sfβgli

Apesar de beta-glicosidases existirem em diversos estados oligoméricos, ainda não é bem conhecida a influência desses estados para a atividade da enzima. Um exemplo conhecido é da enzima de *Exiguobacterium antarcticum* B7, cuja tetramerização resulta em aumento de atividade (Zamphorlin et al., 2016). Considerando a discussão acima, que mostrou que a força iônica induz alterações no estado oligomérico de Sfβgli, a atividade desta enzima nos tampões CF100 e F10 foi determinada.

Foi inicialmente observado que a amostra de Sfβgli em F10 apresentou menor atividade (cerca de 2,5 vezes) quando comparada a amostra em CF100, sugerindo que grandes oligômeros de Sfβgli, possivelmente dodecâmeros, presentes em F10 são menos ativos do que os dímeros preponderantes em CF100. Cabe notar que o pH no ensaio de atividade enzimática é o mesmo nos dois casos e que tanto enzima quanto substrato foram preparados no mesmo tampão, CF100 ou F10, para cada ensaio. No entanto, curiosamente, a atividade da Sfβgli em F10 aumentou com o tempo de armazenagem da amostra a 4 °C, durante dias, indicando um processo lento de ativação da enzima.

Para entender melhor este mecanismo de ativação, foram coletadas curvas de efeito de [S] na atividade da Sfβgli em tampão F10 e em tampão CF100 em diferentes dias após a purificação e incubação da enzima a 4 °C também nestes diferentes tampões. Vale notar que após purificação a mesma amostra de enzima foi dividida em duas alíquotas que agora separadamente passaram pelo processo de troca de tampão (CF100 ou F10) ficando então simultaneamente armazenadas a 4 °C pelo mesmo período de tempo. Após estes diferentes períodos de tempo alíquotas das amostras de enzima foram retiradas e usadas nos experimentos de cinética enzimática feitos a 30 °C.

Observa-se na figura 72 que a Sfβgli em CF100 apresentou cinéticas de Michaelis-Menten muito parecidas para as diferentes amostras de enzima incubadas a 4 °C ao longo de dias. O $V_{\max}$ e K_m oscilaram respectivamente em torno de 39 nmol/min e 2,8 mM, indicando que Sfβgli se manteve estável e inalterada durante a armazenagem a 4 °C. No entanto, a amostra de Sfβgli em F10 (Figura 73) apresentou aumento de atividade ao longo do tempo de incubação a 4 °C devido ao $V_{\max}$, que mais do que dobrou, partindo de 16 para 34 nmol/min, e ao K_m que exibiu uma leve queda, indo de 3,0 mM para 1,9 mM.

Logo, uma primeira informação gerada por este experimento é que a diferença de atividade, mencionada acima, entre os dodecâmeros em F10 e os dímeros em CF100 corresponde preponderantemente a um $V_{\max}$ cerca de 2,5 vezes maior no caso dos dímeros, conforme mostram os parâmetros cinéticos no $t = 0$ de incubação a 4 °C. Vale destacar que o ensaio de atividade das duas amostras foi feito com substrato preparado em CF100, e portanto a força iônica de ambos ensaios torna-se alta, reduzindo efeitos de mudança de atividade produzidas exclusivamente devido a força iônica.

Figura 72. Parâmetros cinéticos da SfbgI selvagem em tampão CF100 em diferentes dias após a purificação incubação a 4°C. No primeiro gráfico acima, curvas de velocidade inicial *versus* concentração de substrato pNPGlu obtidas em diferentes dias após a purificação e incubação da proteína em tampão CF100 a 4 °C. Nos dois gráficos abaixo, valores de K_m e V_{max} em função do tempo de incubação a 4°C da amostra de enzima.

Figura 73. Parâmetros cinéticos da Sfβgli selvagem em tampão F10 em diferentes dias após a purificação incubação a 4°C. No primeiro gráfico acima, curvas de velocidade inicial *versus* concentração de substrato pNPGlu obtidas em diferentes dias após a purificação e incubação da proteína em tampão F10 a 4 °C. Nos dois gráficos abaixo, valores de K_m e V_{max} em função do tempo de incubação a 4°C da amostra de enzima.

Seguindo na análise dos experimentos de cinética enzimática, é curioso notar que após 13 dias a 4 °C a atividade e os parâmetros cinéticos da Sfβgli em tampão F10 tornaram-se próximos daqueles da Sfβgli em tampão CF100, sugerindo que a enzima em F10 passa por um processo lento de ativação que a torna mais parecida com a enzima em CF100, o qual inclui principalmente um aumento de $V_{\max}$ e uma leve diminuição de K_m . Sabendo que as duas amostras contém a mesma concentração de proteína, o aumento de $V_{\max}$ deve refletir um aumento do k_{cat} (Figura 74). Adicionalmente, o processo de ativação é muito mais lento do que a reação de ligação e hidrólise do substrato, logo não há variação da concentração de enzima ativa durante os ensaios a 30 °C e observa-se uma cinética michaeliana clássica.

Considerando que os experimentos de SAXS foram também feitos em tampão F10 e CF100, torna-se natural fazer um paralelo com estes resultados da caracterização cinética. Assim, é possível que o lento processo de ativação de Sfβgli em tampão F10 tenha relação a uma mudança de estado oligomérico da Sfβgli. Desta forma, os grandes oligômeros de Sfβgli, dodecâmeros provavelmente, presente em tampão F10 poderiam estar se dissociando lentamente, formando formas oligoméricas menores com o passar dos dias a 4 °C, tornando-se então mais próximos em termos de atividade e parâmetros cinéticos à Sfβgli dimérica existente em tampão CF100.

Figura 74. Esquema da ativação de SfbgI a 4 °C em tampão F10. E' representa a forma de SfbgI presente em tampão F10 (menor força iônica), possivelmente um dodecâmero. E corresponde a forma mais ativa de SfbgI (provavelmente um dímero). E' é menos ativa que E, conforme expresso por βk_{cat} ($\beta < 1$) e αK_s sendo maior que K_s ($\alpha > 1$). $k_{ativação}$ é a constante de velocidade do processo de ativação, sendo muito menor que o k_{cat} e as constantes de velocidade que compõe K_s .

4.3.5 Análise de SAXS e de atividade enzimática da SfbgI em tampão F10 em diferentes temperaturas

Em uma tentativa de aprofundar o entendimento da correlação entre o estado oligomérico e a atividade da SfbgI, foram coletados perfis de espalhamento de SAXS da enzima em F10 em diferentes temperaturas. Além disso, foram feitos ensaios de atividade da SfbgI sobre o substrato pNPGlu preparado em CF100, em diferentes temperaturas. Desta forma, buscou-se correlacionar possíveis mudanças no estado oligomérico induzidos pela temperatura observadas por SAXS a desvios no plote de Eyring construído com dados de atividade enzimática em diversas temperaturas.

De uma forma geral, observa-se que acima de 40 °C, há um favorecimento ainda maior de estados oligoméricos muito grandes (Figura 75A). Além disso, através de plotes de Krakty, é

possível observar que em temperaturas iguais ou maiores a 45 °C, a suave “curva de sino” presente nas temperaturas de 20 a 30 °C desaparece (Figura 75B), indicando que a enzima não possui formato globular nestas temperaturas, provavelmente devido a sua desnaturação. De fato, a T_m de Sfβgli é de 45 °C. Deste modo, combinando a análise destes dois gráficos nota-se que a desnaturação da Sfβgli em F10 não parece ser precedida da dissociação das unidades do oligômero.

Além disso, analisando-se a atividade enzimática em diferentes temperaturas através do plote de Eyring (Figura 76), nota-se uma diminuição da atividade em temperaturas acima de 40°C, representada na drástica alteração de inclinação da reta. Este efeito também é provocado pela desnaturação da enzima, cuja T_m , como já mencionado, é de 45°C. Além disso, na faixa de 20 a 40 °C a velocidade da reação cresce como esperado para reações químicas e não há mudanças de inclinação na reta, o que indica que não há alterações no mecanismo e etapas limitantes da reação catalisada pela enzima. Já nas curvas de espalhamento não há indícios de modificação clara do grau de oligomerização. Logo, não foi possível estabelecer neste experimento correlações entre o grau de oligomerização e a atividade enzimática de Sfβgli.

Figura 75. Espalhamento por SAXS da Sfβgli em tampão F10 obtido em diferentes temperaturas. **A**, perfis de espalhamento. **B**, plote de Kratky das curvas em diferentes temperaturas.

Figura 76. Plote de Eyring da atividade da Sfβgli obtida em diferentes temperaturas reacionais. Foi usado o substrato pNPGlu preparado em tampão CF100. k corresponde a k_{cat} .

4.3.6 Uso de redes para análise da estrutura quaternária da Sfβgli e perspectivas

Del Sol e O'Meara (2005) observaram que resíduos centrais por *betweenness* da rede são muitas vezes encontrados na interface de ligação entre unidades de oligômeros, indicando que podem ter papel fundamental na interação e na propagação de informação de uma unidade para outra. Além disso, os autores compararam redes produzidas por estruturas monoméricas com redes diméricas de uma mesma proteína, identificando que alguns resíduos centrais do monômero também eram centrais do dímero. Estes resíduos, segundo os autores, muitas vezes são encontrados em *hotspots* da energia de ligação entre unidades, indicando que desempenham papéis importantes na interação entre monômeros do oligômero.

Neste sentido, como observamos que Sfβgli apresenta diferentes graus de oligomerização de acordo com a força iônica, optamos por avaliar como o estado oligomérico pode influenciar na centralidade dos seus resíduos. Para tal, a estrutura dimérica da Sfβgli constituída das cadeias A e B, extraídas da estrutura cristalográfica (5CG0), foi utilizada para modelar a rede de estrutura proteica e determinar os seus resíduos centrais. Além disso, a estrutura tetramérica da Sfβgli contendo as cadeias A, B, E e F também foi analisada. Devido à ausência de uma estrutura monomérica da Sfβgli cristalizada isoladamente, utilizamos como monômero somente a cadeia A da estrutura cristalográfica (5CG0).

De uma forma geral, observa-se que o índice de agregação (C) destas redes é pouco afetado pelo estado oligomérico da enzima, sendo 0,53 para a rede monomérica, 0,52 para a rede dimérica e 0,52 para a rede tetramérica. Além disso, o grau médio dos resíduos também é pouco afetado, sendo 10,3, 10,4 e 10,5 para as redes monomérica, dimérica e tetramérica, respectivamente.

No entanto, observa-se que o parâmetro L aumenta significativamente com o aumento do estado oligomérico da Sfβgli, sendo que seu valor foi de 5,5, 7,9 e 13,0 para as redes monomérica, dimérica e tetramérica, respectivamente. Estes valores são maiores do que o L previsto para redes randômicas ($L_r=3,2$), mas menores do que para redes regulares de igual tamanho, os quais são estimados em 33, 66 e 99, respectivamente². Além disso, é importante observar que o aumento de L com o número de nós ocorre de modo diretamente proporcional em redes regulares, mas não em redes mundo pequeno ou randômicas. De fato, observa-se que

² Para esta estimativa, a rede a curto alcance da Sfβgli foi considerada, sendo esta produzida com *cutoff* para interação entre resíduos tal que o grau médio da rede fosse o mesmo da rede completa (em torno de 10).

em redes mundo pequeno o parâmetro L aumenta proporcionalmente ao logaritmo do número de nós. A análise gráfica deste comportamento mostrou que nas redes estruturais de monômero, dímero e tetrâmero da SfbgI o aumento de L com relação ao tamanho da rede foi mais acentuado do que o previsto em redes mundo pequeno e randômicas (Figura 77). De fato, há sugestão que em um gráfico de L versus $\text{Log}(N)$ os dados da SfbgI seguem o mesmo padrão em curva esperado para rede regular (Figura 77), enquanto os valores para redes mundo pequeno e randômicas apresentam uma tendência linear como previsto pela relação $L \propto \text{Log}(N)$. Isso indica que as redes oligoméricas da SfbgI podem não ser bem descritas pelo modelo mundo pequeno, assumindo um caráter mais regular.

Figura 77. Dependência do caminho mínimo médio da rede (L) com relação ao seu tamanho (N ; número de nós) para diferentes modelos (Regular, Randômico e Mundo Pequeno). O painel à esquerda é uma ampliação da faixa de L 20 do gráfico à direita. Os valores observados na rede oligomérica de SfbgI (●) correspondem ao monômero, dímero e tetrâmero.

Além disso, na figura 78 estão os resíduos centrais por *closeness*, sendo que em vermelho estão os resíduos centrais do monômero (cadeia A de SfbgI) em azul, do dímero (obtidos ao representar conjuntamente as cadeias A e B como uma rede) e em amarelo, os resíduos centrais

do tetrâmero (cadeias A, B, E e F juntas). Nenhum resíduo central por *closeness* no monômero também é central no dímero e no tetrâmero, indicando que o estado oligomérico influencia fortemente esse tipo de centralidade. Observa-se ainda que os resíduos centrais por *closeness* são encontrados no centro de massa da estrutura, enquanto que ao analisar o dímero, eles ficam localizados na interface dimérica, ocorrendo o mesmo com o tetrâmero (Figura 78).

Observando graficamente o processo de perda de centralidade por *closeness* nota-se que ela está distribuída por toda a estrutura da unidade do monômero, tanto na transição para o dímero como para o tetrâmero (Figura 79). Há um padrão repetitivo no processo de perda, indicando que há locais mais drasticamente afetados, talvez coincidentes com as regiões mais centrais na unidade monomérica. Finalmente é curioso notar que na cadeia A a diminuição da centralidade do tipo *closeness* é mais pronunciada do que na cadeia B, que está localizada mais no meio do oligômero.

Por outro lado, ao observar a centralidade por *betweenness*, um fenômeno diferente ocorre (Figura 80). Os resíduos centrais por *betweenness* estão mais distribuídos pela estrutura, conforme já comentado no capítulo 3, e além disso alguns dos resíduos centrais no monômero (E187, P188, S247, N249 e N329) também o são no dímero e no tetrâmero (embora não em todas suas unidades neste último caso; ver abaixo). Dentre estes resíduos, E187, P188 e N329, estão no sítio ativo, sendo que desempenham papéis conhecidos na catálise (Marana *et al.*, 2001; Marana *et al.*, 2003).

Cadeia A - Q34 R97 T139 H142 I183 F185 N186 E187 P188 R189 T245 I246 S247 V328 N329 H330 T398 E399 N400

Cadeia A (parecido com B) - K148 Q150 E151 L152 G153 G154 F155 A156 N157 P158 Y196

Cadeia A e F – nenhum resíduo central

Cadeia B (parecido com E) – D361 S363 W364 L365 L374 P376

Figura 78. Resíduos centrais por *closeness* no monômero (em vermelho), dímero (em azul) e tetrâmero (amarelo) da Sfβgli.

Figura 79. Diferença de centralidade por *closeness* observada para o resíduo *n* analisando a transição de monômero para o dímero da SfbgI (gráfico superior) e de dímero para tetrâmero (gráfico inferior). Neste segundo gráfico a cadeia A está em azul e a B em verde.

Por outro lado, o resíduo N249, que é caracterizado nesta tese, parece ser relevante para a estabilidade e atividade de SfbgI. Outro fenômeno interessante é que na estrutura tetramérica da SfbgI observa-se que as cadeias das extremidades do oligômero apresentam poucos resíduos centrais (somente 3 cada cadeia), sendo que nenhum deles faz parte do sítio ativo.

Deste modo, similarmente ao observado para centralidade por *closeness*, há uma tendência dos resíduos centrais por *betweenness* também migrarem para a interface do dímero e do tetrâmero. Porém, esta tendência não é tão pronunciada como no caso anterior, pois ainda há resíduos centrais no interior das unidades, especialmente no caso do dímero. Finalmente no

caso do tetrâmero nota-se resíduos centrais se estendendo por todo o oligômero, porém sem se interiorizar nas unidades das extremidades.

Assim ao avaliar graficamente as variações na centralidade por *betweeness*, observa-se que alguns resíduos aumentam sua centralidade, enquanto outros diminuem (Figura 81). Porém, o aumento está concentrado em regiões pontuais. Inclusive na transição dímero-tetrâmero, na cadeia B há um número maior de resíduos com aumento na sua centralidade por *betweeness*, possivelmente porque a cadeia B está mais no centro do oligômero.

Amitai e colaboradores (2004) propuseram que os resíduos centrais por *closeness* frequentemente encontram-se no sítio ativo de enzimas. Além disso, como visto no capítulo 3 desta tese, os resíduos centrais por *closeness* e *betweeness* da rede de estrutura proteica parecem desempenhar papéis importantes na catálise da Sfβgli. Com isso, é curioso notar que na cadeia monomérica, esta proposta é verdadeira e se aplica, mas nos outros estados oligoméricos, os resíduos do sítio ativo deixam de ser centrais por *closeness* e especialmente nas unidades dos extremos do tetrâmero por *betweeness*. Esse fenômeno pode estar associado à perda de propriedades Mundo Pequeno nos oligômeros, conforme observado através do aumento do parâmetro L.

Cadeia A – T33 T35 Y38 L93 Y96 L100 P136 L140 Y141 H142 E187 P188 K201 G243 I246 S247 N249 W250 Y275 N329 Y331 I379 E399 G401 S445

Cadeia A (parecido com B) – W44 E51 L100 Y141 L145 K148 G153 G154 A156 N157 P158 S161 E187 P188 Y196 S198 K201 A207 N218 H223 S247 N249 N329 L350

Cadeia A – N112 V114 K148 **Cadeia B** – Q22 L93 N157 E187 P188 I191 Y196 A207 S247 N249 Q301 N329 Y331 F334 P348 D352 T356 S358 A360 D361 S363 W364 L374 P376 S378 Y487 R490 **Cadeia E** – A156 E187 I191 Y196 A207 S247 N249 N329 Y331 E340 L350 T356 S358 A360 D361 S363 W364 L374 Y420 **Cadeia F** – N52 K148 N157

Figura 80. Resíduos centrais por *betweeness* em diferentes estados oligoméricos da *Sfβgli*. Em A, os resíduos centrais do monômero. Em preto estão os resíduos centrais que também são centrais no dímero e no tetrâmero; em vermelho estão os resíduos centrais do monômero, que não são centrais nas outras formas. Em B, os resíduos centrais do dímero. Em roxo, os resíduos centrais comuns entre monômero e dímero, e em azul, resíduos centrais somente do dímero. Em C, os resíduos centrais do tetrâmero. Em verde, os resíduos centrais comuns entre dímero e tetrâmero, e em amarelo, os resíduos centrais do tetrâmero.

Figura 81. Diferenças de centralidade por *betweenness* observadas para o resíduo *n* ao analisar as transições de monômero para dímero (gráfico superior) e monômero para tetrâmero (gráfico inferior). Neste segundo gráfico, em azul está representada a diferença com relação a cadeia A do tetrâmero, e em verde com relação a cadeia B.

A combinação destas observações sugere que o aumento do grau de oligomerização faz com que os resíduos de cada unidade, bem como do oligômero como um todo, fiquem mais distantes entre si em termos de número de contatos. Este “afastamento” se aplica especialmente aos resíduos do sítio ativo, que usualmente eram próximos de todos os demais da estrutura do monômero, segundo a própria definição de centralidade por *closeness*. Este “afastamento” do restante da estrutura proteica também se manifesta na menor presença dos

resíduos do sítio ativo em rotas de contatos curtas entre quaisquer pares de resíduos da estrutura de SfbgI, segundo a própria definição de centralidade por *betweenness*.

Considerando que a centralidade e proximidade do sítio ativo em relação à estrutura da proteína são observadas para enzimas michaelianas e alostéricas (Del Sol e O'Meara, 2005), é possível que esta característica seja relevante para a atividade enzimática. Deste modo, especulamos se a perda de centralidade pelos sítios ativos nos maiores oligômeros, especialmente nas cadeias dos extremos do tetrâmero e outros grandes oligômeros (hexâmeros e dodecâmeros), estaria associada à sua menor atividade enzimática. Neste aspecto, combinando as centralidades por *closeness* e *betweenness*, o dímero talvez ainda exibisse suas unidades suficientemente próximas em termos de rotas curtas de contato, sendo mais ativo. De qualquer modo, deve ser mencionado que caso esta especulação seja correta, não estão sendo invocados efeitos alostéricos entre sítios ativos, pois como já visto anteriormente SfbgI segue uma cinética michaeliana clássica. De qualquer modo, embora esta hipótese seja coerente com os resultados experimentais apresentados neste capítulo, que indicam que formas oligoméricas maiores possuem atividade enzimática menor, ela ainda necessita ser elaborada e testada em maior detalhe, para de fato avaliar se as mudanças nas propriedades físicas e funcionais das diferentes formas oligoméricas da SfbgI estariam associadas a mudanças topológicas da rede estrutural proteica.

4.4 Conclusão

Buscando a caracterização da estrutura quaternária da SfbgI, observamos por SAXS que a enzima se apresenta majoritariamente na forma dimérica em tampão citrato-fosfato 100 mM pH 6,0 (CF100). Além disso, aos resultados apontam a presença de cerca de 10% das proteínas

na forma tetramérica nesta condição. No entanto, ao analisar a Sfβgli em tampão com força iônica mais baixa, fosfato de potássio 10 mM pH 6,0 (F10), observa-se a presença de grandes oligômeros, possivelmente dodecâmeros.

Curiosamente, ao avaliar a atividade dessas amostras, observa-se que amostras da Sfβgli em tampão de baixa força iônica (F10) exibem atividade mais baixa do que amostras em tampão de maior força iônica (CF100). Além disso, a amostra da Sfβgli em baixa força iônica aumenta a sua atividade com o tempo de armazenagem a 4°C, sugerindo que ocorre um lento processo de ativação. Considerando os resultados obtidos por SAXS nestas mesmas condições, uma possibilidade é que esse processo de ativação esteja relacionado a uma mudança no arranjo oligomérico da enzima.

Pensando nos efeitos que arranjos oligoméricos possuem nas redes de estrutura proteica, observamos que a oligomerização de Sfβgli faz com que a centralidade por *closeness* diminua em toda a rede, especialmente no sítio ativo. No dímero e no tetrâmero os resíduos centrais por *closeness* se concentram na interface entre as unidades dos oligômeros. Já ao avaliar a centralidade por *betweenness*, observa-se que em algumas regiões há diminuição, enquanto outras regiões pontuais aumentam a sua centralidade ao analisar monômeros com relação a dímeros e tetrâmeros. Porém, apesar de alguns resíduos permanecem centrais, independentemente do estado oligomérico da Sfβgli, há um processo de perda de centralidade especialmente em resíduos das extremidades do tetrâmero. A combinação destas observações sugere que o aumento do grau de oligomerização de Sfβgli faz com que os resíduos de cada unidade, bem como do oligômero como um todo, fiquem mais distantes entre si em termos de número de contatos. Considerando que a centralidade e proximidade do sítio ativo em relação à

estrutura da proteína são observadas para enzimas michaelianas e alostéricas (Del Sol e O'Meara, 2005; Amitai et al., 2004), é possível que esta característica seja relevante para a atividade enzimática. Deste modo, especulamos se a perda de centralidade pelos sítios ativos nos maiores oligômeros, especialmente nas cadeias dos extremos do tetrâmero e grandes oligômeros (hexâmeros e dodecâmeros, por exemplo), não estaria associada a sua menor atividade enzimática. No entanto, essa ainda é uma hipótese especulativa, e carece de resultados para sua validação.

5

REDES DE ESTRUTURA PROTEICA EM BETA-GLICOSIDASES

5.1 Introdução

5.1.1 Construindo Redes de Estrutura Proteica

O uso de Redes em sistemas biológicos vem crescendo significativamente. Isso porque elas descrevem interações entre entes, o que é de fundamental interesse em sistemas biológicos, pois permite analisar todas as interações de uma forma global.

Outro sistema no qual a análise por redes tem sido uma ferramenta é na estrutura de proteínas. Para isso, estruturas de proteínas são transformadas em redes onde cada resíduo de aminoácido é considerado nó, e aminoácidos que interagem covalentemente ou não covalentemente entre si, são considerados conectados. Diferentemente de grande parte das redes não existe um critério definido para descrever se dois aminoácidos interagem ou não. Vendruscolo e colaboradores consideraram em seu trabalho que, aminoácidos nos quais o C_{α} estão em uma distância menor ou igual a 8,5 Å estariam conectados (Vendruscolo, Manu *et al.*, 2002). Brinda e Vishveshwara desenvolveram um parâmetro de “força” de interação. Nesse parâmetro são consideradas a natureza dos resíduos de aminoácido e o número de contatos que esses aminoácidos fazem em sua cadeia lateral, considerando que para ter um contato, um

átomo da cadeia lateral de um aminoácido deve estar a uma distância menor ou igual a 4,5 Å de um átomo da cadeia lateral de outro aminoácido. Com isso, é criado um número de “força” de interação entre aminoácidos, e é então estabelecido um corte de “força” de interação. Os pares de aminoácidos cuja força de interação está acima do corte são considerados conectados, já os pares cuja força está abaixo do corte são considerados como não conectados (Brinda *et al.*, 2010). O critério que tem sido mais utilizado na literatura foi desenvolvido por Greene e Higman, no qual se considerou que dois resíduos de aminoácidos estariam conectados se algum átomo de um deles estiver a uma distância menor ou igual a 5 Å de algum outro átomo do outro resíduo, que é o limite para a força atrativa de London - van der Waals. No entanto, pouco se sabe sobre como essas diferentes abordagens podem influenciar a rede final.

Um ponto importante das redes é a sua topologia, *ie* a organização geral na qual os nós estão conectados. Neste sentido, Greene e Higman (2003) analisaram conjuntos de 9 tipos de estrutura de proteínas (*trefoil*, *lg fold*, *OB roll*, *plait α/β* , *UP roll*, *lipocalin*, *double wound*, *TIM barrel* e *updown*), totalizando 65 proteínas, e mostraram que a topologia de todas elas, analisadas através da curva de distribuição de k (número de contatos de um nó), aparecia como uma distribuição de Poisson, o que é compatível a uma rede do tipo “Mundo Pequeno”. Neste tipo de rede, parte das conexões entre os resíduos seguem certos padrões, enquanto que outra parte das conexões não seguem padrões claros. São justamente essas conexões, que no caso das Redes de Estrutura Proteica são feitas entre resíduos distantes em estrutura primária, que tornam a rede mais conectada, aproximando trechos distantes da estrutura e encurtando o caminho necessário para ir de um resíduo a outro através de suas conexões.

No entanto, Greene e Higman também observaram que considerando-se somente as interações entre resíduos distantes em estrutura primária, assumiu-se que dois aminoácidos são distantes se estiverem a pelo menos 10 resíduos de distância, a distribuição de k fica mais parecida com uma lei de potências, indicando que o sistema torna-se mais parecido com o modelo “Sem Escala” (Greene e Higman, 2003).

Del Sol e colaboradores (2006) modelaram proteínas como redes considerando que resíduos a uma distância menor do que 5 Å fazem contato. Em seguida, o grupo analisou o parâmetro da via mínima L como índice da centralidade do resíduo na estrutura. Para isso, foi feita uma análise do L_k . Este parâmetro é o L da proteína após deleção computacional de um aminoácido. Assim é possível encontrar resíduos cuja deleção implica em significativo aumento do L_k com relação ao L , que são chamados de resíduos centrais. O grupo analisou, dessa forma, mioglobina, proteína-G, tripsina, hemoglobina, oligossacarídeos fosforilase, domínio de ligação à ligante de receptores nucleares e proteases, encontrando que aminoácidos centrais estão, grande parte das vezes, nos sítios catalíticos e alostéricos das proteínas, e parecem ser importantes na transmissão da informação através da proteína (Del_Sol *et al.*, 2006).

Além do critério de centralidade utilizado por Del Sol, $C_{\Delta Lp}$, outros critérios vêm sendo cada vez mais utilizados para caracterizar proteínas. Dentre estes critérios, temos a centralidade por *closeness*. De uma forma geral, um resíduo central por *closeness* é um resíduo “próximo” a todos os demais resíduos da rede, sendo necessário atravessar pequenas distancias para ir deste resíduo a qualquer outro da estrutura (para mais detalhes, ver capítulo 1). Amitai e colaboradores, analisando um banco de dados de 178 proteínas não redundantes, demonstraram que resíduos centrais por *closeness* são muitas vezes localizados no sítio ativo de

enzimas e que o uso de redes poderia ser uma ferramenta para encontrar resíduos do sítio ativo de proteínas sem a necessidade de alinhamento com outras proteínas (Amitai *et al.*, 2004).

Outro critério bastante utilizado para medir a centralidade dos nós de uma rede é chamado de *betweenness*. Resumidamente, um resíduo central por *betweenness* é um resíduo que muitas vezes é encontrado como integrante das vias mínimas que cruzam a estrutura proteica para ir de um resíduo a outro qualquer (para mais detalhes, ver capítulo 1). Del Sol e colaboradores (2005) têm apontado que, em proteínas, resíduos centrais por *betweenness* são, muitas vezes, *hotspots* de interação entre cadeias proteicas, especialmente em proteínas com estrutura quaternária.

Um critério bastante simples de ser analisado é a centralidade por grau. Neste tipo de centralidade, um resíduo central é aquele que possui um alto número de conexões com sua vizinhança. Brinda e Vishveshwara demonstraram em seu trabalho que, proteínas termofílicas parecem ter um número maior de resíduos altamente conectados do que proteínas mesofílicas, e que a topologia com características “Sem Escala” poderia, portanto, ser importante na estabilidade térmica de proteínas (Brinda e Vishveshwara, 2005). Além disso, Fokas e colaboradores (2016) relatam em seu trabalho que a combinação entre centralidade *closeness* e grau (baixo) podem render importantes informações sobre o resíduo. Isso porque resíduos com baixo grau são, usualmente, mais flexíveis. Enquanto isso, resíduos centrais por *closeness* são mais conservados. Neste sentido, um resíduo que exhibe as duas propriedades seria um resíduo que teria um papel importante, e conservado, na dinâmica da proteína (Fokas *et al.*, 2016).

Desta forma, muitos critérios de centralidade têm sido utilizados na literatura ao caracterizar Redes de Estrutura Proteica. No entanto, existe pouca padronização nos trabalhos

presentes na literatura sobre como construir as redes e como analisá-las. Neste sentido, este capítulo discute as diferentes formas de construção da rede, avaliando se mudanças na construção geram redes parecidas ou significativamente diferentes entre si. Além disso, a centralidade dos resíduos utilizando diferentes critérios será analisada, para avaliar se algum deles melhor se correlaciona a uma dada propriedade proteica.

5.1.2 Redes de Estrutura Proteica e enzimas

Beta-glicosidases são enzimas pertencentes à família 1 das Glicosídeo Hidrolases, a família mais abundante de enzimas existente na natureza. Estas enzimas possuem enovelamento do tipo barril (α/β)₈, uma estrutura bastante comum entre enzimas (Wierenga, 2001). Desta forma, estas enzimas são bastante representativas, e seu uso como modelo neste capítulo pode revelar propriedades comuns a outros tipos de enzimas. No entanto, as Redes de Estrutura Proteica de beta-glicosidases ainda são pouco exploradas pela literatura.

Avaliando a Rede de Estrutura Proteica de diferentes beta-glicosidases, Ikegami, em seu doutoramento, observou que os resíduos centrais por $C_{\Delta Lp}$ nem sempre são conservados, exibindo um certo grau de variação do resíduo (Ikegami, 2013). No entanto, os sítios de centralidade de beta-glicosidases parecem bastante conservados, sendo que seus resíduos centrais aparecem muitas vezes em posições análogas em estrutura terciária. Isso sugere que, apesar de resíduos centrais por $C_{\Delta Lp}$ nem sempre serem conservados, eles podem desempenhar importantes papéis na comunicação ao longo da estrutura proteica e que, portanto, a organização da estrutura de forma que esses resíduos sejam centrais pode ser essencial para manutenção de suas propriedades.

Além disso, Fokas e colaboradores, avaliando a centralidade de proteínas constituintes de *Saccharomyces cerevisiae*, observaram que os resíduos centrais por *closeness*, *betweeness* e *grau* possuem uma menor taxa de evolução do que os demais resíduos. Isso indica que pode haver algum tipo de pressão evolutiva que torna esses resíduos mais conservados do que os demais resíduos não centrais (Fokas *et al.*, 2016).

Nesse sentido, o objetivo deste capítulo é avaliar a conservação dos resíduos centrais (por diferentes critérios) de diferentes beta-glicosidases, o que pode indicar a presença de pressão evolutiva para manter estes resíduos, indicando a sua possível correlação com propriedades físicas importantes deste tipo de enzimas.

5.2 Materiais e Métodos

5.2.1 Construção da Rede de Estrutura de Sfβgli e cálculo da centralidade $C_{\Delta Lp}$

A construção da rede da estrutura de Sfβgli e a determinação da centralidade dos seus resíduos através do parâmetro $C_{\Delta Lp}$ foram feitos segundo descrito no capítulo 1.

5.2.2 Produção das redes utilizando o aplicativo *CMView* 1.1.1

Para produzir as redes de estrutura proteica considerando a distância entre os resíduos como critério de estabelecimento de um contato, foi utilizado o programa *CMView* 1.1.1. Para isso, o programa foi instalado segundo o manual do fabricante, sendo que seu funcionamento é associado ao programa *Pymol*. Dessa forma, ao iniciar o programa *CMView* 1.1.1, o *Pymol* funciona simultaneamente, permitindo a visualização da rede de estrutura tridimensionalmente (Figura 82A). As estruturas em formato PDB da Sfβgli e de outras beta-glicosidases foram abertas

utilizando o programa *CMView* e as redes foram produzidas considerando diferentes critérios. Em todas as estruturas estudadas, somente a cadeia A foi analisada quando oligômeros estavam presentes nos arquivos PDB. Na construção da rede foram utilizados como critérios de contato entre dois resíduos a distância entre o par de átomos mais próximos e também a distância entre os seus carbonos-alfa. Foram considerados conectados os resíduos que tivessem átomos a uma distância de pelo menos “d”, chamada de distância corte. Foram testadas distâncias corte de 2 a 7 Å. Além disso, também se testou produzir a rede considerando que dois resíduos estão conectados se tiverem carbonos-alfa a uma distância corte de 8,5 Å. Após a produção pelo programa *CMView*, o mapa de contatos das diferentes proteínas em estudo (Figura 82B) foi exportado e posteriormente aberto no programa *Cytoscape* para uma análise mais detalhada da rede, segundo o item 5.2.3.

Figura 82. Produção da rede de estrutura proteica da Sfβgli considerando a distância entre resíduos como critério para formação de contatos. Em A, a estrutura representada como rede visualizada no programa *Pymol*. Em B, uma representação do mapa de contatos gerado pelo programa *CMView*.

5.2.3 Análise das redes de estrutura proteica utilizando o programa *Cytoscape*

Após os procedimentos descritos no item 5.2.2, a rede de estrutura proteica foi caracterizada utilizando o programa *Cytoscape* 3.2.0. Para isso, a rede foi analisada utilizando o comando *Analyze Network*. Após esta análise, uma planilha foi gerada contendo a centralidade de cada um dos resíduos por *closeness*, *betweenness* e grau, bem como os parâmetros gerais da rede (L_p e C_p), segundo descrito em Resultados e Discussão.

5.3 Resultados e Discussão

5.3.1 Construção da rede de estrutura proteica utilizando diferentes critérios de conexão entre nós

Para iniciar esta análise, diferentes redes de estrutura proteica da SfbgI foram produzidas no programa CMView considerando como critério de contato entre resíduos uma distância de pelo menos 5 Å entre quaisquer átomos, de acordo com o proposto por Greene e Higman (2003), ou então a uma distância de pelo menos 8,5 Å entre carbonos-alfa, de acordo com o proposto por Vendruscolo e colaboradores (Vendruscolo *et al.*, 2001). As redes construídas através destes critérios estão representadas em forma circular na figura 83. Na figura observa-se semelhança entre as redes construídas segundo os diferentes critérios de contato. Além disso, observa-se que na rede construída utilizando carbonos-alfa como referência, o número de conexões totais da rede é um pouco maior, sendo 2805 conexões para a rede produzida considerando todos os átomos, e 2901 conexões quando considera-se apenas carbonos-alfa para estabelecer conexões. O grau médio da rede produzida tendo como

referência os carbonos-alfa também é um pouco maior, sendo de 11,9 graus, enquanto que a rede produzida com todos os átomos esse valor é de 11,5. Apesar disso, na rede produzida através dos carbonos-alfa, observa-se que os resíduos estão mais conectados com a sua vizinhança próxima e menos conectados com os resíduos distantes, pois o índice de agregação C_p desta é maior do que da rede produzida por todos os átomos (0,53 e 0,48, respectivamente), e o caminho mínimo médio da rede também é um pouco maior (5,1 e 4,8, respectivamente). Isso demonstra que as redes produzidas das duas formas é bastante semelhante, mas que a rede produzida a partir de carbonos-alfa resulta em mais conexões locais.

Figura 83. Comparação entre redes de estrutura proteica produzidas a partir do modelo cristalográfico da Sfβgli (5CG0) utilizando diferentes definições de contatos entre resíduos. Em A, contatos são estabelecidos se dois resíduos tiverem quaisquer átomos a 5 Å. Em B, contatos são estabelecidos se carbonos-alfa estão a 8,5Å. Os resíduos foram coloridos de acordo com a sua centralidade por *closeness*, sendo os mais vermelhos, mais centrais.

Adicionalmente, temos que na rede construída segundo o critério de contato baseado em carbonos-alfa a 8,5 Å os resíduos centrais possuem uma distribuição muito semelhante àquela da rede onde contato foi definido como quaisquer átomos a 5 Å. Na tabela 11, é possível observar que 15 resíduos aparecem como centrais por *closeness* quando a distância de corte é

entre átomos a pelo menos 5 Å e 19 resíduos aparecem como centrais ao considerar uma distância de corte de pelo menos 8,5 Å entre carbonos-alfa. É importante notar que estas duas listas de resíduos centrais possuem muitas semelhanças, sendo que 10 deles são comuns. Além disso, os segmentos de estrutura primária representados nestas duas listas são sobreponíveis, exceto pelo W444, resíduo formador da plataforma hidrofóbica do subsítio -1 do sítio ativo de Sfβgli, que aparece apenas na primeira listagem.

Tabela 11. Resíduos centrais por *closeness* da Rede de Estrutura Proteica produzida a partir da estrutura cristalográfica da Sfβgli (5CG0) utilizando diferentes critérios de contato. Ao lado, o elemento de estrutura secundária em que eles aparecem ou estão muito próximos.

<i>Critério de contato</i>	
<i>Quaisquer átomos a 5,0 Å</i>	<i>Cα a 8,5 Å</i>
R97 (β2)	R97 (β2)
Y141 (β3)	T139 (β3)
H142 (β3)	H142 (β3)
W143 (β3)	I183 (β4)
I183 (β4)	F185 (β4)
F185 (β4)	N186 (β4)
N186 (β4)	E187 (β4)
E187 (α4)	P188 (α4)
I191 (α4)	I244 (β5)
T245 (β5)	T245 (β5)
N329 (β6)	I246 (β5)
Y331 (β6)	S247 (β5)
T398 (β7)	G327 (β6)
E399 (β7)	V328 (β6)
W444 (β8)	N329 (β6)
	H330 (β6)
	T398 (β7)
	E399 (β7)
	N400 (β7)

Elementos de estrutura secundária, α, alfa-hélice; β, fita beta, estão numerados de acordo com sua ordem relativa no (β/α)₈ barril

Adicionalmente temos que as redes estruturais de Sfβgli produzidas utilizando distâncias de corte entre quaisquer átomos ou entre carbonos-alfa para estabelecer contatos entre resíduos são semelhantes. Esse resultado também foi observado por Cheng e colaboradores (2016), que comparando redes estruturais proteicas produzidas de diferentes formas concluíram que são igualmente apropriados os critérios para estabelecer contatos entre resíduos baseados em átomos a pelo menos 5 Å (de acordo com o utilizado nesta tese), carbonos-alfa a pelo menos 10 Å e centro de massa da cadeia lateral a pelo menos 10 Å. Contudo Fokas e colaboradores (2016) indicam que obtiveram resultados melhores utilizando como contatos apenas as ligações de hidrogênio e interações de van der Waals, apesar de também terem tido bons resultados utilizando a distância de corte como critério.

Tendo assumido com base na discussão acima que distâncias de corte são critérios apropriados para definição de contatos em estruturas proteicas e construção de redes estruturais, deve ser mencionado que Greene e Higman (2003) discutem que 5 Å entre átomos é o limite máximo no qual pode haver interação de van der Waals. No entanto, este critério é bastante “generoso” no sentido que pode produzir conexões entre resíduo que não necessariamente estariam interagindo. Desta forma, testes foram feitos utilizando diferentes distâncias para corte, para avaliar como esta afetaria as propriedades da rede.

Foram construídas Redes de Estrutura Proteica da Sfβgli considerando que resíduos estariam em contato se tivessem ao menos um par de seus átomo a uma distância de 2, 3, 4, 5, 6 e 7 Å (Figura 84).

Figura 84. Redes de estrutura proteica da Sfβgli produzidas a partir da estrutura cristalográfica 5CG0 utilizando diferentes distâncias de corte para definição de contatos entre resíduos. Em A, corte de 2 Å; Em B, corte de 3 Å. Em C, corte de 4 Å. Em D, corte de 5 Å; Em E, corte de 6 Å; Em F, corte de 7 Å.

Ao comparar os parâmetros topológicos das redes (Tabela 12), temos que distâncias de corte maiores geram redes com mais conexões, o que faz com que os resíduos tenham maior grau em média. Na rede estrutural com limite de 2 Å, não há ligações a longo alcance, permanecendo naturalmente apenas a estrutura covalente da Sfβgli devido à distância de corte muito curta, como pode ser observado na figura 84. Deste modo, nesta rede estrutural cada resíduo tem apenas dois contatos, que correspondem aos seus vizinhos na estrutura covalente da proteína. Diferentemente das demais, a rede estrutural a 2 Å apresenta propriedades de uma rede regular (ver capítulo 1), com L_p extremamente longo (163), afinal o único caminho até outros nós é através da estrutura covalente, e C_p igual a zero, pois não há trios de resíduos em contato, pré-requisito para formação de *clusters*. Por outro lado, um aumento na distância de corte gera um maior parâmetro de C_p , indicando que nestas redes encontra-se mais agregados, provavelmente devido a um maior número de conexões na rede. Nota-se que o C_p aumenta drasticamente a partir da distância de corte de 3 Å, convergindo acima de 4 Å para um valor estável próximo a 0,5, o qual é muito maior que C_p de uma rede randômica (Figura 85). Por fim, ao comparar os valores de L_p das redes, temos que redes com cortes muito baixos possuem grandes valores de L_p , indicando que é necessário atravessar muitos contatos para ir de um nó a outro através de conexões. No entanto, com distâncias de corte acima de 4 Å, o valor de L_p torna-se muito baixo, chegando a 3,7 quando considera-se contatos de até 7 Å (Figura 85). Isso sugere que a rede torna-se mais conectada, sendo necessário atravessar um menor número de contatos para ir de um nó a outro na rede, em média. Portanto, em distâncias de corte acima de 4 Å já estão estabelecidas as características de “Mundo Pequeno” da rede estrutural proteica de Sfβgli, com C_p muito maior que aqueles de redes randômicas e regulares e L_p semelhante àquele

de uma rede randômica, mas muito menor que uma rede regular (Watts e Strogatz, 1998). Em suma, de uma forma geral, observa-se que todas as redes geradas com limites maiores do que 4 Å tem características semelhantes.

Tabela 12. Parâmetros das redes de estrutura proteica da Sfβgli produzidas com diferentes distâncias de corte como critério para estabelecer conexões.

Corte (Å)	Grau médio	L_p	C_p
2	2,0	163	0
3	3,2	9,5	0,06
4	8,8	5,4	0,45
5	11,5	4,8	0,48
6	14,6	4,2	0,50
7	20,2	3,7	0,51

Outro fato importante é que os resíduos centrais das redes estruturais de Sfβgli produzidas com diferentes distâncias de corte variam muito pouco (Tabela 13). Observa-se que o número de resíduos significativamente centrais por *closeness* em diferentes cortes fica em torno de 15. Esta estabilização no número e listagem dos resíduos centrais por *closeness* é coerente com a estabilização do parâmetro L_p a partir de 4 Å (Figura 85).

Finalmente, a convergência dos parâmetros L_p e C_p para valores estáveis compatíveis como uma rede de “mundo pequeno” e a invariabilidade dos resíduos centrais por *closeness* em distâncias de corte acima de 4 Å mostram que o critério de contato em 5 Å sugerido por Greene e Higman (2003) e usado nesta tese se mostrou robusto e apropriado para gerar uma rede estrutural de Sfβgli com propriedades compatíveis com as de Redes de Estrutura Proteica descritas na literatura.

Figura 85. Parâmetros das redes estruturais da Sfβgli em função da distância de corte para contato entre resíduos.

Tabela 13. Relação de resíduos centrais por *closeness* em redes estruturais de Sfβgli produzidas com diferentes distâncias de corte como critério de estabelecimento de conexões entre resíduos.

Distância de corte				
3 Å	4 Å	5 Å	6 Å	7 Å
G32	R97	R97	Q39	R97
T33	Y141	Y141	R97	Y141
T35	H142	H142	Y141	H142
Y96	W143	W143	H142	W143
R97	N186	I183	F185	F185
F98	E187	F185	N186	N186
W143	I191	N186	E187	E187
F185	T245	E187	I246	R189
N186	V328	I191	N329	T245
E187	N329	T245	Y331	I246
P188	H330	N329	T398	N329
V248	E399	Y331	E399	Y331
N329	N400	T398	W444	T398
H330	W444	E399		E399
Y331		W444		W444
N400				

5.3.2 Análise comparativa da rede estrutural de Sfβgli

Para comparações a Rede de Estrutura Proteica obtida para a Sfβgli foi produzida com o programa Cytoscape utilizando a cadeia A da estrutura cristalográfica (5CG0). Como critério para contatos considerou-se que resíduos interagem se tiveram átomos a uma distância corte de 5 Å, segundo discutido no item anterior. A partir desta rede foram calculados os parâmetros C_p e L_p no programa *Cytoscape*, sendo que os valores encontrados foram 0,48 e 4,8, respectivamente. Estes valores estão de acordo com aqueles obtidos anteriormente, $C_p = 0,49$ e $L_p = 4,8$, utilizando um modelo de homologia de Sfβgli e um programa construído pelo Prof. Guilherme M. Arantes, como descrito no item 1.2.1 (Ikegami, 2013). Estes parâmetros refletem a característica de “mundo pequeno” da rede estrutural de Sfβgli, pois C_p é muito maior que o coeficiente de *clustering* para uma rede aleatória de tamanho comparável (0,02), enquanto L_p é comparável ao caminho mínimo médio da rede aleatória (2,6) (Watts e Strogatz, 1998). A representação da rede gerada em forma circular pelo *CMView/Cytoscape* (Figura 86A) também permite notar visualmente que está possui característica “Mundo Pequeno”, onde os resíduos fazem contatos de longo alcance com outros mais distantes em estrutura primária, mas que se aproximam tridimensionalmente (Figura 86B), e também mantém contatos de curto alcance com resíduos próximos em estrutura primária (Figura 86C). Estes resultados estão de acordo com os resultados obtidos para outras proteínas descritos na literatura (Greene e Higman, 2003; Bagler e Sinha, 2005; Aftabuddin e Kundu, 2007).

Figura 86. Características “Mundo Pequeno” das redes de estrutura proteica de Sfβgli. Em A, a representação circular da rede, e o seu respectivo histograma de distribuição de graus dos nós. Em B, a representação circular da subrede de contatos a longo alcance, com o seu respectivo histograma de distribuição de graus. Em C, representação da subrede de contatos a curto alcance, com seu respectivo histograma de distribuição de graus.

Dissociando a rede estrutural de Sfβgli em duas subredes, uma considerando apenas contatos entre resíduos próximos (curto alcance) e outra entre resíduos ao menos 10 posições distantes em estrutura primária (longo alcance), temos que com relação aos parâmetros C_p e L_p , a rede a longo alcance possui C_p de 0,147 e L_p de 5,71, o que indica que esta rede possui uma maior aleatoriedade do que a rede completa e exibe um grau de agregação muito baixo. Por outro lado, ao analisar a rede a curto alcance, temos que o valor de C_p é de 0,610 e o L_p é de 41,2, o que indica que esta rede possui um caráter mais regular do que a rede completa, exibindo um caminho mínimo médio acentuadamente mais alto. Desta forma, temos que a agregação da rede de estrutura de Sfβgli é uma propriedade cuja principal componente está nos contatos entre resíduos próximos na estrutura primária, enquanto que o caminho mínimo médio curto ocorre principalmente graças às conexões de longo alcance.

Quando analisamos a rede da estrutura de Sfβgli tomando todos os contatos na distância corte de 5 Å, temos que o número médio do grau de conexões dos resíduos é 11,5. Além disso, nota-se que a distribuição de graus dos resíduos na rede completa gera uma distribuição de Poisson, conforme o ajuste na figura 86A. Isso indica mais uma vez que a rede da estrutura de Sfβgli possui propriedade de “Mundo Pequeno”. Já ao analisar as subredes, temos que o número médio de contatos dos resíduos é 5,3 na rede a longo alcance e 6,2 na rede a curto alcance, o que indica que em média os resíduos de Sfβgli fazem um número ligeiramente maior de contatos com resíduos próximos em estrutura primária do que com resíduos distantes. Porém, temos que as distribuições de contatos destas subredes não se ajustam em nenhuma das equações propostas pelos modelos, distribuição de Poisson e lei das potências. Dessa forma, analisando visualmente as distribuições nota-se que no caso da subrede de longo alcance há

maior frequência de resíduos com 10 ou mais contatos do que na subrede de curto alcance. Enfim, na subrede de longo alcance é mais provável encontrar resíduos muito conectados do que na subrede de curto alcance. Esta característica aproxima a subrede de longo alcance de uma rede “Sem Escala” (ver capítulo 1), porém como comentando acima nenhum dos modelos descreve perfeitamente esta subrede, que possui características únicas. Greene e Higman (2003), em seu trabalho, sugerem que a distribuição a longo alcance pode ser ajustada pela lei semilogarítmica ($y = \lambda a^x$), e desta forma a rede seria do tipo “Escala Única”, um modelo de redes que compartilha algumas propriedades da rede “Sem Escala”. O fato dos resíduos da rede estrutural proteica não poderem estabelecer contatos indefinidamente, pois têm uma área superficial limitada, seria determinante para que a rede de longo alcance não se ajustasse perfeitamente ao modelo “Sem Escala” e tendesse ao “Escala Única” (Greene e Higman, 2003). Já Morita e Takano (2009) indicam que esta distribuição de graus semilogarítmica está associada a redes fractais, e utilizando cálculos das dimensões da rede, concluem que as Redes de Estrutura Proteica fazem parte dos *Clusters* de Percolação, contido dentro da classe das Redes Fractais.

5.3.3 Comparação de diversos tipos de centralidade na rede estrutural de Sfβgli

Diversos grupos têm apontado que resíduos centrais da rede estrutural desempenham papéis importantes em propriedades e características de proteínas. No entanto, existem várias formas de calcular a centralidade de um resíduo, conforme já descrito no item 1.1.5, capítulo 1. Desta forma, para comparar os diferentes tipos de resíduos centrais na Rede Estrutural da Sfβgli, a centralidade dos resíduos foi calculada utilizando diferentes parâmetros. A centralidade $C_{\Delta Lp}$

(Del_Sol *et al.*, 2006) foi calculada com colaboração do Prof. Dr Guilherme M. Arantes, em um programa desenhado para essa finalidade (Ikegami, 2013). Já a centralidade em Grau, *Closeness* e *Betweenness* foram calculadas através do programa *Cytoscape*, utilizando a ferramenta de análise de rede. Os resultados obtidos estão representados na figura 87, onde a cor e tamanho dos círculos, que correspondem aos resíduos, são atribuídos de acordo com a sua centralidade, sendo os mais vermelhos e maiores aqueles mais centrais. Além disso, a tabela 14 lista os resíduos mais centrais de acordo com os parâmetros analisados. Estas listagens incluem os resíduos cujo parâmetro de centralidade foi maior do que 3 desvios padrão da média, no caso dos parâmetros de centralidade $C_{\Delta Lp}$ e em *betweenness*. Curiosamente, ao analisar a centralidade em *closeness* e grau, observou-se que nenhum resíduo teve um valor de centralidade 3 desvios padrão maior do que a média. Desta forma, na tabela 5 estão representados os resíduos com valores de centralidade 2 desvios-padrão maiores do que a média para a centralidade em *closeness* e grau.

Na figura 87 e tabela 14 é possível observar que os resíduos centrais calculados por diferentes algoritmos diferem entre si, mas há casos de resíduos que constam de três listagens, como R97 e Y331.

De uma forma geral, o parâmetro de centralidade por grau é semelhante entre os resíduos da rede estrutural e desta forma existem poucos nós cuja centralidade por grau destaca-se em relação à média. Ao analisar o histograma com os valores de centralidade por grau dos resíduos da Sfβgli (Figura 88), observa-se que este parâmetro fica na maioria dos casos contido em um intervalo de ± 3 desvios padrão da centralidade média, seguindo uma distribuição normal. Isto indica que não existem diferenças significativas entre eles, somente

oscilações ao acaso. Isso faz com que na Figura 87 tenhamos resíduos centrais (pontos vermelhos) regularmente espalhados ao longo de toda a rede estrutural.

Figura 87. Representação dos resíduos centrais da Rede de Estrutura Proteica da SfbgII. Em A, os resíduos centrais por grau. Em B, os resíduos centrais por *closeness*. Em C, os resíduos centrais por *betweenness*. Em D, os resíduos centrais determinados segundo C_{ALP} . As cores e tamanho dos resíduos estão de acordo com a centralidade, exceto em D que não foi produzido automaticamente, onde todos os resíduos centrais estão pintados da mesma cor (vermelho) e com o mesmo tamanho.

Tabela 14. Relação de resíduos centrais na Rede de Estrutura da Sfβgli segundo diferentes parâmetros.

<i>Centralidade</i>			
<i>C_{ΔLP}</i> *	<i>Grau</i> #	<i>Closeness</i> #	<i>Betweenness</i> *
R97	Y96	R97	Y96
T245	Y196	Y141	R97
S247	Y230	H142	Y141
N249	W272	W143	H142
F251	R477	I183	E187
Y331	Y483	F185	I191
F334		N186	N249
S358		E187	W272
K366		I191	Y331
E399		T245	W371
Y420		N329	W444
		Y331	F460
		T398	
		E399	
		W444	

*Resíduos cujo parâmetro de centralidade difere em relação à média em pelo menos 3 desvios padrão. #Resíduos cujo parâmetro de centralidade diferente em relação à média em pelo menos 2 desvios padrão. Os resíduos centrais por *C_{ΔLP}* foram já identificados na tese de C.M. Ikegami (Ikegami, 2013) e também reportados em Souza e colaboradores (2016).

Ao analisar a centralidade por *closeness*, um fenômeno semelhante ocorre, sendo que a distribuição dos valores de centralidade por *closeness* também se assemelha a uma curva normal (Figura 88). Porém, a localização dos resíduos centrais por *closeness* mostra tendência de concentração em pontos específicos da rede estrutural de Sfβgli (Figura 87), especificamente ocorrendo em regiões com muitas conexões de longo alcance que os interconectam. Deste modo, embora não haja valores de centralidade que se destaquem, os resíduos centrais em *closeness* não são uma ocorrência do acaso.

Já a distribuição da centralidade por *betweenness* na rede estrutural de Sfβgli apresenta características diferentes, desviando significativamente de uma curva normal (Figura 88). Observa-se que há grande número de resíduos com baixa centralidade, o qual diminui com o

aumento da centralidade, mas ainda assim existe uma população de resíduos com centralidade maior do que 3 desvios padrão. Esse mesmo tipo de distribuição também ocorre para a centralidade $C_{\Delta Lp}$. Deste modo, há resíduos significativamente diferentes da população pelos critérios centralidade em *betweenness* e $C_{\Delta Lp}$, os quais tendem a se concentrar em pontos específicos da rede estrutural de Sfβgli (Figura 87). Notadamente os resíduos centrais por $C_{\Delta Lp}$ tendem a ocorrer nos pontos onde concentram-se os resíduos centrais por *closeness*, havendo também notável sobreposição nos pontos de concentração dos resíduos centrais por *closeness* e *betweenness*.

Ao mapear na estrutura secundária de Sfβgli as regiões que concentram os resíduos centrais por *closeness*, observa-se que elas estão presentes em fitas β do barril $(\beta/\alpha)_8$. Nota-se, por exemplo, que as folhas β_2 , β_3 , β_4 , β_5 , β_6 , β_7 e β_8 contêm resíduos centrais por *closeness* (Figura 89, onde resíduos mais centrais estão indicados por nós em vermelho).

Figura 88. Distribuição da centralidade dos nós da rede estrutural de Sfβgli utilizando diferentes critérios. Em A, Centralidade por *closeness*. Em B, por grau. Em C, por *betweenness*. Em D, centralidade $C_{\Delta LP}$. Mais, número de desvios padrão maior que 3,5.

Figura 89. Representação da estrutura secundária e da rede estrutural da Sfβgli produzidos a partir da sua estrutura cristalográfica (5CG0). Retângulos vermelhos simbolizam regiões de alfa-hélice e setas azuis, regiões de fitas beta. Na rede circular, os nós estão coloridos de acordo com a sua centralidade, sendo que quanto mais vermelhos, mais centrais. Em A, estão representados os resíduos centrais por grau. Em B, os resíduos centrais por *closeness*. Em C, os resíduos centrais por *betweenness*. Em D, os resíduos centrais por $C_{\Delta LP}$.

Complementando este mapeamento, os resíduos centrais por *closeness* foram também localizados na estrutura terciária de Sfβgli (Figura 90). Nota-se que os resíduos com alta centralidade normalmente ocorrem na região final de fitas-beta (3 esferas azuis) e nas alças (10

esferas cinzas) imediatamente posteriores a estas fitas-beta. Apenas um resíduo central ocorre em alfa-hélice (esfera vermelha). Tendência semelhante é notada analisando a centralidade por *betweenness* (Figuras 89 e 90), observando-se poucos resíduos centrais em alfa-hélices (2; esferas vermelhas na figura 90), mas um maior número em alças (8 esferas cinza) e fitas-beta (2; esferas azuis). Similarmente, os resíduos centrais por $C_{\Delta Lp}$ ficam mais concentrados em fitas beta (6 esferas azuis) do que em alças (3 esferas cinza) ou alfa hélice (1 esfera vermelha), preponderantemente na metade C-terminal de Sfβgli. Apenas os resíduos com alto grau parecem estar mais concentrados em alfa-hélices (5 esferas vermelhas).

Dessa forma, dada a preponderante presença de resíduos centrais por $C_{\Delta Lp}$, *closeness* e *betweenness*, podemos concluir que o topo do beta-barril (extremidade C-terminal das fitas-beta e sua conexão com as alças) da estrutura de Sfβgli possui pontos com alta proximidade dos demais resíduos da proteína, estando ligado a eles por rotas que envolvem um reduzido número de conexões (Figuras 88 e 89).

Amitai e colaboradores (2004), utilizando parâmetro de centralidade por *closeness*, observaram que o sítio ativo de enzimas é enriquecido em resíduos centrais e que esse fenômeno é independente do tipo de dobramento das enzimas. Isso indicaria que o sítio ativo é uma região próxima, em termos de número de contato, a todos os demais resíduos da estrutura da proteína. Szalay e Csermely (2013) observaram através da comparação de diferentes tipos de centralidade na rede de estrutura proteica da Met-tRNA sintase, que os resíduos centrais por *closeness* também estão concentrados na região do sítio ativo. Analisando a tabela 14, observa-se que, de fato, grande parte dos resíduos centrais por *closeness* também estão na região do sítio ativo da Sfβgli, sendo que os resíduos R97, H142, E187, N329, Y331, E399 e W444 são

conhecidos por participarem da catálise da reação de hidrólise de forma direta ou indireta (Marana *et al.*, 2001; Marana *et al.*, 2003; Marana, 2006). Os demais resíduos centrais por *closeness* estão próximos na estrutura primária aos resíduos acima mencionados e, portanto, também próximos do sítio ativo.

Figura 90. Representação dos resíduos centrais na estrutura da Sfβgli. Em A, os resíduos centrais por *closeness*. Em B, os resíduos centrais *betweenness*. Em C, os resíduos centrais por $C_{\Delta LP}$. Em D, os resíduos centrais por grau. Nas representações de estrutura, os resíduos em vermelho estão contidos em regiões de alfa-hélice; os resíduos em azul, em fitas beta; os resíduos em cinza, em alças.

Finalmente, é importante notar que em enzimas com enovelamento barril $(\beta/\alpha)_8$ o sítio ativo sempre ocorre na região C terminal de fitas beta ou na região imediatamente posterior (Wierenga, 2001), onde também observamos a concentração dos resíduos centrais por *closeness* e $C_{\Delta Lp}$ em Sf β gli (Figuras 89 e 90). Desta forma, temos que a presença de resíduos altamente centrais que deixam as extremidades das fitas-beta e alças a um curto número de contatos de toda a estrutura proteica pode ter sido evolutivamente determinante para ocorrência de sítios ativos nessa região dos barris $(\beta/\alpha)_8$.

Além da co-localização com sítios ativos, analisando também a centralidade por *closeness*, alguns autores propõem que ela estaria inversamente associada com a flexibilidade local do resíduo. Atilgan e colaboradores (2004) compararam o inverso do parâmetro *closeness* (correspondente ao caminho mínimo médio do resíduo para todos os outros da rede) com o *B-factor* para cada resíduo da estrutura cristalográfica de proteínas, encontrando que esses dois parâmetros são relacionados. Analisando a Sf β gli da mesma forma (Figura 91), observa-se a mesma tendência, indicando que resíduos centrais por *closeness* possuem uma flexibilidade local menor.

Em suma, os resíduos centrais por *closeness* tendem a se concentrar no sítio ativo e pontos menos flexíveis da estrutura de Sf β gli

Figura 91. Relação entre o parâmetro de centralidade por *closeness* e o *B-factor* de cada um dos resíduos de Sfβgli. Em A, gráficos de *B-factor* e do inverso de *closeness* em função do resíduo. Em B, um gráfico relacionando o *B-factor* com o parâmetro de *closeness* dos resíduos. *B-factors* foram extraídos da estrutura cristalográfica 5CG0 (Tamaki et al., 2016)

5.3.4 Análise de rede de beta-glicosidases da família GH1

Para avaliar se a rede obtida para a Sfβgli tem semelhança às redes de estrutura proteica de outras beta-glicosidases GH1, foram selecionadas 21 estruturas cristalográficas de beta-glicosidases. Para esta análise, foram consideradas beta-glicosidases de diferentes organismos, Bacteria, Archea e Eucariotos.

Produzindo as redes para as beta-glicosidases, é possível observar que os parâmetros C_p , L_p e número médio de conexões são bastante próximos entre todas elas (Tabela 15).

Tabela 15. Parâmetros da rede de estrutura proteica de diferentes beta-glicosidases GH1.

Código PDB	Organismo	Grau médio	C_p	L_p	Nós	Conexões
3wq8	<i>Pyrococcus furiosus</i>	11,5	0,487	4,63	448	2568
1cbg	<i>Trifolium repens</i>	11,4	0,486	4,78	490	2787
1gon	<i>Streptomyces sp.</i>	11,4	0,482	4,63	448	2545
1qox	<i>Bacillus circulans</i>	11,8	0,486	4,60	449	2650
1v02	<i>Sorghum bicolor</i>	11,6	0,492	4,73	484	2808
1vff	<i>Pyrococcus horikoshii</i>	11,5	0,486	4,53	423	2440
2e3z	<i>Phanerochaete chrysosporium</i>	11,8	0,491	4,61	460	2703
2jfe	<i>Homo sapiens</i>	11,5	0,488	4,70	459	2636
2o9p	<i>Paenibacillus polymyxa</i> (BgB)	11,5	0,487	4,60	445	2557
2wbg	<i>Thermotoga maritima</i>	11,6	0,485	4,62	443	2566
3ahx	<i>Clostridium cellulovorans</i>	11,4	0,486	4,61	444	2542
3ahy	<i>Trichoderma reesei</i>	11,6	0,485	4,63	465	2696
3cmj	Desconhecido (metagenoma)	11,5	0,486	4,62	441	2536
3vif	<i>Neotermes koshunensis</i>	11,7	0,486	4,66	473	2776
3w53	<i>Micrococcus antarcticus</i>	11,5	0,482	4,67	469	2686
3wh5	Desconhecido (metagenoma)	11,5	0,482	4,58	440	2537
3wq8	<i>Pyrococcus furiosus</i>	11,5	0,487	4,63	448	2568
4ptx	<i>Halothermothrix orenii</i> (BgA)	11,4	0,482	4,63	445	2545
4qlk	<i>Oryza sativa</i>	11,7	0,491	4,66	472	2765
5cg0	<i>Spodoptera frugiperda</i>	11,5	0,481	4,79	488	2805
5dt5	<i>Exiguobacterium antarcticum</i> B7	11,5	0,485	4,60	444	2555
5foo	<i>Streptococcus pyogenes</i>	11,5	0,487	4,67	463	2667

Além disso, na figura 92 é possível observar que a distribuição de contatos da rede estrutural é semelhante entre as diferentes beta-glicosidases, sendo que em todas elas há a presença de contatos a longo alcance formando um “triângulo invertido”, que indica as conexões do sítio ativo entre os resíduos R97, E187 e E399 (a numeração utilizada está de acordo com a Sfβgli). Para efeitos comparativos, na figura 92 também está ilustrada a rede estrutural da triosefosfato isomerase (PDB: 1TIM), que também é um barril (β/α)₈, mas naturalmente possui sequência primária significativamente diferente de beta-glicosidases GH1.

Observa-se que a distribuição dos resíduos centrais por *closeness* é semelhante entre TIM e as beta-glicosidases GH1, reproduzindo-se então a tendência da extremidade do barril-beta ser próximo ao restante de toda a proteína. Esta similaridade sugere que esta é uma propriedade de barris $(\beta/\alpha)_8$.

5.3.4.1. Centralidade por grau

Para avaliar o alinhamento dos resíduos centrais por grau de diferentes beta glicosidases, inicialmente a centralidade por grau de cada resíduo foi calculada utilizando o programa *Cytoscape*. A partir disso, resíduos com valores de grau maiores em pelo menos 2 desvios-padrão da média foram determinados para cada uma das beta-glicosidases. Na figura 93 está representado um alinhamento das sequências das beta-glicosidases analisadas com os resíduos centrais por grau marcados com fundo colorido. Os resíduos marcados com fundo preto têm valores de centralidade por grau maiores do que 3 desvios-padrão, os com fundo cinza escuro tem valores maiores do que 2,5 desvios-padrão e com fundo cinza, valores maiores do que 2 desvios padrão da média.

1CBG

1GON

1QOX

1V02

1VFF

2E3Z

2JFE

2O9P

2WBG

3AHX

3AHY

3CMJ

Figura 92. Representação circular das redes de estrutura proteica de beta-glicosidases GH1. Em vermelho estão representados os resíduos centrais por *closeness*. 1CBG corresponde ao código da beta-glicosidase de *Trifolium repens*; 1GON, de *Streptomyces sp.*; 1QOX - *Bacillus circulans*; 1V02 - *Sorghum bicolor*; 1VFF - *Pyrococcus horikoshii*; 2E3Z - *Phanerochaete chrysosporium*; 2JFE - *Homo sapiens*; 2O9P - *Paenibacillus polymyxa* (BgB); 2WBG - *Thermotoga maritima*; 3AHX - *Clostridium cellulovorans*; 3AHY - *Trichoderma reesei*; 3CMJ - Desconhecido (metagenoma); 3VIF - *Neotermes koshunensis*; 3W53 - *Micrococcus antarcticus*; 3WH5 - Desconhecido (metagenoma); 3WQ8 - *Pyrococcus furiosus*; 4PTX - *Halothermothrix orenii* (BgA); 4QLK - *Oryza sativa*; 5CG0 - *Spodoptera frugiperda*; 5DT5 - *Exiguobacterium antarcticum* B7; 5FOO - *Streptococcus pyogenes*. 1TIM corresponde a Triosefosfato Isomerase de *Gallus gallus*.

De uma forma geral, observa-se que existem 2 resíduos altamente conectados, E41 e Y196, que ocorrem na maior parte das beta-glicosidases analisadas. Comparando as diferentes sequências, observa-se um alto grau de conservação desses resíduos. Além disso, observa-se que o único resíduo com grau de conexão 3 desvios acima da média é o Y196, que em algumas beta-glicosidases assume esse valor muito acima da média. O papel destes resíduo nas

propriedades das beta-glicosidasases não é conhecido. Outros resíduos centrais por grau, no entanto, estão distribuídos de forma esparsa ao longo do alinhamento, sendo que as suas posições não são nitidamente conservadas. Na análise de centralidade por grau em Sfβgli (Tabela 14 e Figura 88) já foi sugerida uma distribuição randômica. Interessantemente a maioria dos resíduos centrais por grau são resíduos com cadeia lateral volumosa, como triptofano e tirosina, possivelmente porque tem maior superfície para fazer interações.

[illegible]

3wq8	85	R I FPKPTFDVKVDEKEEENIIISVDVPSTIKELEKIANMEALEH Y RKIYSDWKERGKTFFILNL Y HWPLPLWIHDPIAVRKLGPDAAPAGWLDE-KT VV	193
3wh5	84	RIVPLGAG-----PINQAGLDH Y SRMVDALLGAGLQPFFVTLY HWDLPPQLDR-----LGWGSR-AT AT	141
lgon	96	R QPTGRG-----PALQGGLDF Y RRLADELLAKGIQPVATLY HWDLPPQELNA-----GGWPER-A T AE	153
3w53	123	R VVPGGR-----TVNPEGLDF Y SRLVDELLENGILPWLTLY HWDLPPQALEER-----GGWTNR-E TS Y	179
2o9p	92	R IMPAAG-----IINEEGLLF Y EHLLEDIELLAGLPMTLY HWDLPPQWIEDE-----GGGWQR-E T IQ	148
3cmj	104	R IQPDSSR-----QINQRGLDF Y RRLVEGLHKRDILPMATLY HWDLPPQWVEDE-----GGWLSR-E S AS	161
3ahx	84	R IFPKGFG-----EINQKGIGF Y RDLLIDELLKNDIEPATIY HWDLPPQLQDT-----GGWANP-Q V AD	141
lqx	83	R VLPGGTG-----EVNRAGLDY Y HRLVEDELLEANGIEPFCTLY HWDLPPQALQDQ-----GGWSLR-I T ID	140
3ahy	87	R I PEGGRG-----DAVNQAGIDHY VKFVDDLLDAGITFPFITLY HWDLPEGHQRY-----GGGLNTEFL PL	148
2e3z	87	R I IPKGGRS-----DPNVGAGIKH Y RTLIEELVKEGITPFVTL Y HWDLPPQALDDRY-----GGWLNKEEA IQ	148
lv02	155	R L LPKGTALA-----GGINEKKREV Y NKLIDLLENGIEPYITI F HWDTPQAALVDAY-----GGFLDE-R I IK	215
3vif	91	R VLPEGHG-----NIYNQDGIDY Y NNLINELLANGIEPMVTM Y HWDLPPQALQDL-----GGWPNI-V LA K	149
lcbg	98	R VLPKGLKLS-----GGVNEREGINY Y NNLINEVLANGIPYVTF HWDVPPQALEDEY-----RG FLGR-N I VD	158
2wbg	106	R ILPEGTG-----RVNQKGLDF Y NRIIDITLLEKGITPFVTL Y HWDLPPFALQLK-----GGWANR-E I AD	163
2jfe	82	R LLPDGTG-----GFINGKGDY Y NKIIDDLLKNGVTPITVL Y HFDFLPQTLEDQ-----GGWLSE-A I IE	140
5foo	81	R L IDDFEQ-----ATINPDGLAY Y NRVIDACLANGIRPVNL HHFDLPIALYQAY-----GGWSES-HVV D	140
5dt5	104	R I FPSK-----QQNFPEGMAF Y KTALRLQEEGIKPAVTL Y HWDLPMWAHEE-----GGWWRN-D S VD	160
5cg0	84	R I LPTGMA-----NEVNPAGIAF Y NNYIDEMLKYNI TPLITLY HWDLPPQKLQEL-----GGFANP-L I SD	142
4qlk	98	R IFPDGE-----GRVNGEGVAY Y NNLIYNLLQKITPPYNNLY HYDLPLALEKKY-----GGWLNA-K M AD	156
4ptx	85	R ILPEGK-----GRVNQKGLDF Y KRIVDNLLKANIRPMITLY HWDLPPQALQDK-----GGWTNR-D T AK	142
lvff	75	R L FPEN-----KFNEDEFAMK Y RLIVDLITLRTGLTPLVTL HHFTSPWFMMKK-----GGGLRE-E NL K	141

3wq8	184	E F VKFAAFVAYHLLDLVDM	W T M N E P N V Y V N Q G	I N L A S G P P G P F L	S F E A A E K A K F N L I Q A H I G A Y D A I K E Y S E	266
3wh5	142	V F A E Y A D I V V R Q L G D R V T H	W A T L N E P W C S A M L G Y	Y L G - V H A P G H T	D L K R G L E A S H N L L L G H G L A V Q A M R A A P Q F	224
1gon	154	R F A E Y A A T A A D A L G D R V K T	W T T L N E P W C S A L L G Y	G S G - V H A P G R T	D P V A A L R A A H H L N L G H G L A V Q A L R D R L P A D	236
3w53	180	K F L E Y A E T H V E K L G D R V K H	W T T N E P L C S S L L G Y	A A G - E H A P G R Q	E P Q A A L A A V H Q L L H A L G A T A R L R E L G A	260
2o9p	149	H F K T Y A S V I M D R F G E R I N	W N T I N E P Y C A S I L G Y	S T G - E H A P G H E	N W R E A F T A A H I L M C H G I A S N L H K E G L T	230
3cmj	162	R F A E Y T H A L V A A L G D Q I P L	W T V T N E P M V T V W A G Y	H M G - L F A P G L K	D P T L G G R V A H H L L L S H G Q A L Q A F R A L S P A G	244
3ahx	142	Y Y V D Y A N L L F R E F G D R V K T	W I T N E P W V A S Y L G Y	A L G - V H A P G I K	D M K M A L A A H N I L L S H F K A V K A Y R E L E Q D	223
1gox	141	A F A E Y A E L M F K E L G G K I K	W I T N E P W C M A L S N Y I	A G - V H A P G N K	D L Q A I D V S H H L L V H G R A V T L F R E L G I S	222
3ahy	149	D F A E Y A M V F R A - L P K V R N	W I T N E P L C S A I P G Y	G S G - T F A P G R Q	S T S E P T V T G V H N I L V A G R A V K A Y R D E F K A S G D G Q I G I V L G	232
2e3z	149	D E T N Y A K L C F E S F G D L V Q N	W I T N E P W V I S V M G Y	G N G - I F A P G H V	S N T E P W I V S H H I L A H A H A V K L Y R D E F K E K Q G	233
1v02	216	D Y T D F A K C F E K F G K T V K N	W L T N E P E T F C S V S	Y G T - V L A P G R C S P G V S C A V P T G N S	L S E P Y I V A H N L L R A H A E T V D I Y N K Y H K G A D	311
3vif	150	Y S Y E N A R V L K F N F G D R V K L	W T L T N E P L T F M D G Y	A S E I - G M A P S I N	T P G I D G Y L A A H Y I Q L H A R I Y H L Y D Q E F R A E Q G	235
1cbg	159	D F R Y A E L C F K F G D R V K H	W I T N E P W G V S N A Y	A Y E I - T F A P G R C S	D W L K N C T G D G S G R E P Y L A A H Y I L A A A A R Y I K T K Y Q A S Q N	255
2wbg	164	W F A E Y S R V L F E N F G D R V K N	W I T L N E P W V A I V G	H L Y G - V H A P G M R	D I Y V A F R A V H N L L R A H A R A V K V F R E T V K D	245
2jfe	141	S F D K Y A Q P C S F T G D R V K V	W I T I N E A N V L S V M S	Y D L G - M P P G P I	H F G T G Q Y A A H N L I K A H A R S W H S Y D S L F R K K Q G	226
5foo	141	L F V A S K V C F E Q F G D R V K D	W F V H N E P M V V E G S	Y L G - F H Y P A I V D	G K K A V Q V A Y N L A L T A K V I A Y Y R R G P A E L S	225
5dt5	161	W F D F S K V F E E L G I V D S	W I T N E P W C A G E L S	Y H L G - Q H A P G H T D	M N E A V R A V H M L L S H G K A V E M L K E G F N S	243
5cg0	143	W F E D Y A R V V F E N F G D R V K M	F I T N E P R E I C F E G	Y G S A - T K A P I L N A	T A M G A Y L C A K N L V T A H A K A Y Y L Y D R E F R P V Q	228
4qlk	157	L F T E Y A D C P K F T G N R V K H	W F T N Q P R I V A L L G	Y D Q G - T N P K R C T K C A	A G G N S A T E P Y I V A H N F L L S H A A A V A Y R T K Y Q A A Q	249
4ptx	143	Y F A E Y A R M F E E F N G L V D	W T T N E P W V A F E G	H A F G - N H A P G T K D	K L Q V A H H L L L S H G M A V D I F R E E D L	224
1vff	132	H W E K Y L E K V A E L - L E K V K L	V V T N E P W V M M G L	I T A - Y W P P F I R	S P F K A F K V A A N L L L S A T A Y E L L H G K	209

Figura 93. Alinhamento de sequências de beta-glicosidases e comparação dos resíduos centrais por grau.

5.3.4.2. Conservação dos Resíduos Centrais por *betweeness*

Na figura 94 está ilustrado um alinhamento das sequências das beta-glicosidases comparando seus resíduos centrais pelo parâmetro de *betweeness*. Os resíduos com valor de centralidade maior do que 3 desvios padrão da média estão destacados com fundo preto. Já os resíduos com centralidade maior do que 2,5 desvios padrão da média estão destacados com fundo cinza escuro e maior do que 2 desvios estão destacados com fundo cinza.

É possível observar que alguns resíduos centrais por *betweeness* são comuns entre as beta-glicosidases GH1, enquanto outros aparecem de forma esparsa no alinhamento. Dentre os resíduos centrais encontrados podemos destacar 8 regiões altamente centrais.

Uma delas é a região da fita beta 2 e início da alfa-hélice 2, que contém os resíduos Y96, R97 e W102, que são resíduos altamente conservados e que são centrais por *betweeness* em quase todas as beta-glicosidases analisadas. O resíduo R97 desempenha papel fundamental na catálise da Sf β gli, atuando como modulador do pK_a do resíduo catalítico E399 (Marana *et al.*, 2003).

Uma segunda região altamente central é a região logo após a fita beta 3, que contém dois resíduos frequentemente centrais, os resíduos Y141 e H142. Estes dois resíduos são bastante conservados em beta-glicosidases, sendo que o resíduo H142 faz parte do sítio ativo e atua na ligação do glicone do substrato (Rozenfeld, 2006; Tamaki *et al.*, 2016).

Outra região altamente central está contida na alça que antecede a alfa-hélice 3, sendo que um resíduo de triptofano frequentemente ocorre na posição 155. No entanto, apesar do razoável grau de conservação, Sf β gli possui outro resíduo nessa posição, uma fenilalanina, que

não aparece como resíduo central nesse caso. Além disso, duas outras beta-glicosidases analisadas, as de *Pyrococcus horikoshii* e de *Thricoderma reesei*, também possuem substituição deste triptofano por uma leucina e uma fenilalanina, respectivamente. Nestes dois casos, a posição também não aparece como central. Até o momento, a importância deste triptofano não foi caracterizada para beta-glicosidases GH1.

Uma região bastante interessante de alta centralidade aparece no final da fita-beta 4 e dentro da alfa-hélice subsequente. Nesta região, o resíduo E187 aparece como altamente central e também conservado entre as diferentes de beta-glicosidase. Este ácido glutâmico é um dos resíduos catalíticos das beta-glicosidases, atuando como doador de prótons na catálise ácido-base (Marana *et al.*, 2001). Além disso, diversos resíduos aparecem como centrais nessa região, de forma esparsa e não conservada, principalmente na alfa-hélice 4. Esta região é conhecida por conter resíduos que formam o subsítio de ligação do aglicone do substrato (Mendonca e Marana, 2008). Sabendo que beta-glicosidases podem assumir especificidades bastante diversificadas pelo aglicone (Esen, 1993), essa variabilidade dos resíduos centrais desta alfa-hélice pode estar relacionada a esta diversidade na especificidade pelo aglicone.

Outra região que possui alta centralidade, mas os resíduos não são conservados e estão esparsos é alfa-hélice 5. No entanto, não há até o momento estudos do papel dos resíduos centrais desta região nas propriedades das beta-glicosidases.

Já na fita-beta 6, temos uma região com diversos resíduos centrais conservados. Um deles é o resíduo Y331, que faz parte do sítio ativo da Sf β gli e outras beta-glicosidases, atuando na modulação do pK_a do resíduo catalítico E399 (Marana *et al.*, 2003). A região da fita beta 7, na qual o próprio resíduo E399 aparece como altamente central, também é conservada na maior

parte das beta-glicosidases GH1. Este resíduo desempenha papel catalítico nas beta-glicosidases, promovendo o ataque nucleofílico necessário para a catálise (Marana *et al.*, 2001).

Por fim, temos que a região da fita-beta 8 e de uma outra fita beta subsequente apresenta dois resíduos altamente centrais bastante conservados em beta-glicosidases. Um deles é o W444, que forma a plataforma hidrofóbica do subsítio -1 para ligação do aglicone do substrato, tendo fundamental importância na catálise. Outro resíduo central e conservado é o F460, que já foi caracterizado e é importante na catálise promovida por Sf β gli mesmo estando distante do sítio ativo (Mendonça e Marana, 2011).

3wq8 3 -----KFPKNFMFGYSISGFOEMGLPGS-EVESDWWVWVHDKENIASGLVSGDLPENGPAYWHLYKQDHDIAEKLGMDCIRGGIEWA 84
3wh5 4 -----ERFPADFVWGAATAAYQIEGAVREDGRGVSIWDTFSSH-TPG---KIADGTDGDVACDSYHRYGEDIGLLNALGMNAYRFSIAWP 83
1gon 15 -----ALTFFPEGFLWGSATASQIEGAAEDGRTPSIWDTYAR-TPG---RVRNGDGDVATDHYHWRREDVALMAELGLGAYRFSIAWP 95
3w53 36 ----MNLHSQ-----KFAWPKFELWGSATAAQIEGAGHSYKGEDSVWDADFAR-KEG---AIAGGENLEVAVDHYHRYREDVQLMRELGLDSYRFSIAWP 122
2o9p 10 -----NTFIFPATFMWGTSTSSYQIEGGTDEGGRTPSIWDTFCQ-IPG---KVIGDGCDDVACDHFHFKEDVQLMKQLGLFLAYRFSIAWP 91
3cmj 24 -----KFPFEGFLWGAATSSYQIEGAWNEDGKGESIWDRTFR-IPG---KIKNGDSGDVACDHYHRYEQDLDLMRQLGLKAYRFSIAWP 103
3ahx 1 -----MEK-----LRFPKDFIFGTATAAYQIEGAYKEDEKGESIWDRTFSH-IPG---NVAKMHNGDIACDHYHRYKEDVQLLKSLGKAYRFSIAWP 83
1qox 1 -----SIHMFPSDFKWGVATAAYQIEGAYNEDGRGMSIWDTFAH-TPG---KVKNGDGNVACDSYHRYVEEDVQLLKDLGVKAYRFSIAWP 82
3ahy 8 -----MLPKDFQWGFATAAYQIEGAVDQDGRGSPSIWDTFCA-QPG---KIADGSSGVTAACDSYNRNRAEDIALKSLGAKSYRFSIAWP 86
2e3z 6 -----AAKLKPSFVWGVATAAYQIEGSPDKDGRREPSIWDTFCQ-APG---KIADGSSGDVATDSYNRNREDVQLLKSYGVKAYRFSIAWP 86
1v02 64 ---RLSPWEI---PRRDWFPFSLFGAATSAYQIEGAWNEDGKGPSIWDTHCFHNFP---WIVDRSNGDVAADSYHMYAEDVRLLEKMGMDAYRFSIAWP 154
3vif 7 -----DTV-----YTFFDEFKLGAATASQIEGAWDENGKGNPIWDTLTTHHPD---YVVDGATGDIADDSYHLYKEDVKILKELGAQYRFSIAWP 90
1cbg 1 FKPLPISFDDFSDLNRSFCFAPGFFVFGTASSAFQYEGAAFEDGKGSPSIWDTFTHKYPE---KIKDRTNGDVAIDYHRYKEDIGIMKMDNLDAYRFSIAWP 97
2wbg 25 -----VKKFPFEGFLWGVATAAYQIEGSPDADGAGMSIWDTHFSHTP-G---NVKNGDGTGDVACDHYHRYKEDVQLLKGLGVKAYRFSIAWP 105
2jfe 3 -----FPAGFGWAAATAAYQVEGGWDADGKGCPVWDTFTTHQGG---RVFKNQTDGVACGSSYTLWEEDLKKIKQLGLTAYRFSIAWP 81
5foo 1 -----MLAPPKEFWGGATSGPQSEG---RFAQKHRLNIDYWEYEEEPD---LFYDYVGPDTASDAYHQIESDLTLLASLGHNSYRFSIAWP 80
5dt5 25 -----KFAPNFVFGTATSSYQIEGAGHDEGGRTPSIWDTFCD-TDG---KVFEKHNGDVAACDHYHRYKEDVQLLKGLGVKAYRFSIAWP 103
5cg0 2 -----QRRFPDDFLFGTATASQIEGAWDEGKGGENIWDYVMVHTPE---VIRDLNNGDIAADSYHNYKRDVEMMRELGLDAYSFSLWA 83
4qlk 10 -----NWLGLSRAAFPKRFVFGTVTSAYQVEGMAASGGRGSPSIWDAFAH-TPG---NVAGNQNGDVAATDQYHRYKEDVNLMSKSLNFDAYRFSIAWP 97
4ptx 5 -----IFPEDFIWGAATSSYQIEGAFNEDGKGESIWDRTFSH-TPG---KIENGDTGDIACDHYHLYREDIELMKIEGIRSYRFSIAWP 84
1vff 1 -----MPLKFPFEMFLFGTATSSHQIEGNNR---WNDWWYEQ-IGK---LPYRS-GKACNHWELYRDDIQLMSTLGYNAYRFSIAWP 74

3wq8 85 RIFPKPTFDVKVDVEKEEGNISVDVPSTIKELEKIANMEALEHYRKIYSOWKERGKTFILNLYHWPLPLWIHDPIAVRKLGPDAAPAGILDE-KTVV 183
3wh5 84 RIVPLGAG-----PINQAGLDHYSRMVDALLGAGLPQFVTIYHMDLPQPLEDR-----LGWGSR-ATAT 141
1gon 96 RIQPTGRG-----PALQKGLDFYRRRLADELLAKGIQPVATYHMDLPQLENA-----GGIPER-ATAE 153
3w53 123 RVVPGGR-----TVNPEGLDYFYSRLVDLELLENGILPLTLTYHMDLPQALEER-----GGITNR-ETSY 179
2o9p 92 RIMPAAG-----IINEEGLLFYEHLLDEIEAGLIPMLTYHMDLPQWIEDE-----GGITQR-ETIQ 148
3cmj 104 RIQPDSSR-----QINQRGLDFYRRRLVEGLHKRDLIPMATYHMDLPQWVEDE-----GGILSR-ESAS 161
3ahx 84 RIFPKFGF-----EINQKGIQFYRDLIDLELKNDEIPAITYHMDLPQKLQDI-----GGWANP-QVAD 141
1qox 83 RVLPGGTG-----EVNRAGLDYHRLVDLELLANGIEPFTLYHMDLPQALQDQ-----GGWSSR-ITID 140
3ahy 87 RIIPFEGGR-----DAVNQAGIDHYVYKFDVLLDAGITPFTITLYHMDLPGLHQRQY-----GGLNRTPEPL 148
2e3z 87 RIIPKGGRS-----DPVNGAGIKHYRTLIEELVKEGITPFVTLYHMDLPQALDDRY-----GGWLNKEEAQ 148
1v02 155 RILPKGTLA-----GGINEKRVEYNNKLIDLLLENGIEPYITIFHMDLPQALVDAY-----GGELDE-RIIK 215
3vif 91 RVLPEGHD-----NIVNQDGIYYNNLIDLELLANGIEPMTYHMDLPQALQDL-----GGWPNL-VIAK 149
1cbg 98 RVLPGKGLS-----GGVNRREGIYNNNLEIVLANGMQPVYTLHMDVPQALEDEY-----RGFLGR-NIVD 158
2wbg 106 RILPEGTG-----RVNQKGLDFYNNRIIDTLLLEKGIPTFVTLYHMDLPFALQLK-----GGWANR-EIAD 163
2jfe 82 RLLPDGTT-----GFINQKGIYYNNKIIDLLKNGVTPIVTYHMDLPQTLLEDO-----GGWLSE-AIEE 140
5foo 81 RLIDDFEQ-----ATINPDGLAYYNNRVIDACLANGIRPVNIHFDLPALYQAY-----GGWESK-HVVD 140
5dt5 104 RIFPSK-----RFGMAFYKTLATRLQEEGKPAVTYHMDLPMWAHE-----GGWVNR-DSVD 160
5cg0 98 RILPTGMA-----NEVNPAAGIAYNNYIDEMLYKNTIPLITYHMDLPQKLQEL-----GGFANP-LISD 142
4qlk 84 RIFPDGE-----GRVNPQEGVAYNNLNYLLQKGITPYVNIYHMDLPALAEKKY-----GGWLNK-KMAD 156
4ptx 85 RILPEGK-----GRVNPQGLDFYKRLVDNLLKANIRPMITYHMDLPQALQDK-----GGWITN-DTAK 142
1vff 75 RLFPEEN-----KFNEDAFMKYREIIDLTLRGITPLVTLHFTSPFLWFMKK-----GGFLRE-ENLK 131

3wq8 184 EFVKFAAFVAYHLLDLDVDMSTNDPNVYNGQYINLASGFPFGFL-----SFEAAEKAKFNLIQAHIGAYDAIKEYSE-----KSVGVIYAF 266
3wh5 142 VFAEYADIVVRQLGDRVTHWATLNEPWCASALGYLYG--VHAPGT-----DLKRGLLEASHNLLGHGLAVQAMRAAPQP--LQIGIVLNL 224
1gon 180 RFAEYAAIAADALGDRVKWTTEPWCASAFGLYGG--VHAPGT-----DPVAALRAAHNLLGHGLAVQALRDRLPAD--AQCSVTLNI 236
3w53 154 KFLYEAYTVHEKLGDRVKHWTTNEPLCSSLIGYAG--EHAPGRQ-----EPQAALAAVHHQLAHGLATARLRELGA--EHIGITLNI 260
2o9p 149 HFKTYASVIMDRFGERINWNNTINEPYCASILGYGTG--EHAPGHE-----NWREAFTHAHLTMCHGIASNLHKEKGLT--GKIGITLNI 230
3cmj 162 RFAEYTHALVAALGDQIPLVVTHNEPMVYVWAGYHMG--LFAPGLK-----DPTLGGRVVHLLLSHGQALQAFRALSPAG--SQMGITLNI 244
3ahx 142 YVVDYANLLRFEGDRVKTWVTHNEPMVASYLGYALG--VHAPGIK-----DMKMALLAHHNLLSHFKAVKAYRELEQD--GQIGITLNI 223
1qox 141 AFAEYAEMLFKELGGKIKQWITNEPWCASALPGLYGG--TFAPGRQ-----DQLAIDVSHLLLVAGHRAVTLFRELGIS--GEITLAPNT 222
3ahy 149 DFENYARVMFRA-LPKVRNWTNEPWCASALPGLYGG--TFAPGRQ-----STSEPTVGHNLVVAHGRAVKAYRDDFKPASGDGQIGIVLNG 233
2e3z 149 DETNYAKLCFESFGDLVQNWITNEPMVLSVMGYGNG--IFAPGHV-----SNTPEWIVSHHILAHAAVAVKLYRDEFKEQG--GQIGITLNI 233
1v02 216 DYTDFAKVCFEKFGKTVKNWLTNEPTEFSYSYGTG--VLAPGRCSPGVSCAVPTGNSLSEPIYVAHNLRAHAETVDIYNKYHKGAD--GRIGLALNV 311
3vif 150 YSENYARVLKFNFGDRVKLTLTNEPLTFMDGYASEI--GMAPSIN-----TPGIGD--LAATVIHAHARIYHLYDQEFRAEQG--GKVGISLNI 235
1cbg 159 DFRDYAELCFKEFGDRVKHWITNEPMVGSMAAYAG--TFAPGRCSDWLKLNCTGGDSGREPYLAHYQLAHAAARLYKTKYQASQ--GIIGITLVS 255
2wbg 164 WFAEYSRVLFEFNGDRVKNWITNEPMVVAIVGHLYG--VHAPGMR-----DIYVAFRAVHLLRAHARAVKVFRETVDK--GKIGIVFNN 245
2jfe 141 SFDKYAQCFSTFGDRVKWITNEANVLSVMSYDLG--MFPPGIP-----HFGTGGY--AAHNLKAKHARSWSYDLSFRKKQK--GMVSLSLFA 226
5foo 141 LFVAFSKVCFEQFGDRVKWVHNEPMVWVGSYLMG--FHPAIVD-----GKKAVQVAYNLAATAKVIQAYRRGPAELS--DGRIGITLNI 225
5dt5 161 WFLDFAVCFEELDGIYDSWVTHNEPWCAGFLSYHMG--QHAPGHTD-----MNEAVRAVHMLLSHGKAVLEMLKEFNS--ATPIGITLNI 243
5cg0 143 WFDYARVVFENFGDRVKMITNEPRELFEYGSA--TKAPILNA-----TAMGAYLCAKNLVTAHAKAYLYDREFRPVQ--GGCGGITISV 228
4qlk 157 LFTFYADFCKFTFGRNVKHWITNEPRIALGLYDQ--TNPPKRCCTKCA---AGGNSATEPIYVAHNFLLSHAAVARYRTKYQAAQ--QGKVGIVLDF 249
4ptx 143 YFAEYARLMFEFNGLVLDLVVTHNEPMVVAIEGHAFG--NHAPGTD-----FKTALQVAHLLLSHGMAVDIFREEDL--PBEIGITLNI 224
1vff 132 HWEKYIEKVAEI-LEKVKLVATNEPMVVMGYLTA--YWPPIR-----SPFKAFKVAANLLKAHAIAVELLHGK-----FKVGIVKNI 209

Figura 94. Alinhamento de sequências de diferentes beta-glicosidases GH1 com destaque para os resíduos centrais por *betweeness*. Os resíduos com valor de centralidade maior do que 3 desvios padrão da média estão destacados com fundo preto. Já os resíduos com centralidade maior do que 2,5 desvios padrão da média estão destacados com fundo cinza e maior do que 2 desvios estão destacados com fundo cinza. Regiões de alfa hélices estão em vermelho e fitas beta em azul.

5.3.4.3. *Conservação dos Resíduos Centrais por closeness*

Outro tipo de centralidade cuja conservação foi analisada em beta-glicosidases GH1 foi a centralidade dos resíduos por *closeness* (Figura 95). Deve ser notado que nenhuma beta-glicosidase apresentou resíduos com valores de centralidade por *closeness* maiores do que 3 desvios padrão da média. Assim, quando consideramos resíduos com valores de centralidade entre 2 e 3 desvios padrão temos a presença de 7 regiões com resíduos centrais. Os resíduos R97, Y142, H143, N186, E187, N329, Y331, E399 e W444, que aparecem como altamente centrais em praticamente todas as beta-glicosidases GH1 e que são altamente conservados, desempenham papéis conhecidos na atividade destas enzimas. Por outro lado, o resíduo T245 aparece como central na maioria das beta-glicosidases analisadas, mas não é conservado e não tem papel conhecido. No entanto, no capítulo 1 nota-se que a mutação T245A faz com que Sf β gli seja expressa em corpos de inclusão, o que sugere que esse resíduo central que ocupa o *core* proteico pode ter papel importante na sua estrutura.

Interessantemente os resíduos centrais por *closeness* frequentemente se apresentam em fitas beta ou logo após as fitas-beta, como discutido no item 5.3.2. Observando as beta-glicosidases, das 8 fitas beta principais pertencentes ao barril central, 7 possuem resíduos centrais ou dentro da própria fita ou logo em seguida a ela. Além disso, outras regiões destas proteínas praticamente não apresentam resíduos centrais por *closeness*, indicando que a região do barril-beta e as alças imediatamente posteriores são regiões que estão a poucos contatos de quaisquer outros resíduos da proteína. É interessante destacar que esta região do barril-beta e as alças imediatamente posteriores é exatamente onde ocorre o sítio ativo, não só em beta-glicosidases GH1, mas em todas as enzimas que possuem estrutura do tipo barril (β/α)₈.

(Wierenga, 2001). Conforme já comentado anteriormente, a posição do sítio ativo nestas enzimas pode estar relacionada à maior proximidade desta região com o restante da estrutura proteica, indicando que esta proximidade pode ser uma propriedade importante para o funcionamento do sítio ativo.

3wq8 3 -----KFPKNFMFGYSWSGFQ FEMGLPGS-EVESDWWVWVHDKENIASGLVSGDLPENGPAWYHLYKQDHDIAEKLGMDCIRGGIEWA 84
3wh5 4 -----ERFPADFLVNGAA T AAY Q IEGAVREDGRGVSIWDTFSH-TPG---KIADGTTGDVACDSYHRYGEDIGLLNALGMNAYRFSIAWP 83
1gon 15 -----ALTFFPEGLWNGSA T ASY Q IEGAAADGRTPSIWDTYAR-TPG---RVRNGDTGDVADTHYHRRWEDVALMAELGLGAYRFSLAWP 95
3w53 36 -----MNLHSQ-----KFAWPKFELWNGSA T AAA Q IEGAGHSYKGEDSVWDAFAR-KEG---AIAGGENLEVAVDHYHRYEDVQLMRELGLDSYRFSISWA 122
2o9p 10 -----NTFIFPATFMWGTS T SSY Q IEGGTDEGGRTPSIWDTFQ-IPG---KVIIGDCGDVACDHFHFKEDVQLMKQLGLFHYRFSVAMP 91
3cmj 24 -----KKFPEGLWNGAA T SSY Q IEGAWNEDGKGESIWDRTFR-IPG---KIKNGSDGDVACDHYHRYEQDLDMRQLGLKTYRFSIAWA 103
3ahx 1 -----MEK-----LRFPKDFIFGTA T AAY Q IEGAYKEDEKGESIWDRTFSH-IPG---NVAKMNGDIADCHYHRYKEDVQLKSLGKISYRFSIAWP 83
1qox 1 -----SIHMFPSPDFKWGVA T AAY Q IEGAYNEDGRGMSIWDRTFAH-TPG---KVKNGDNGVACDSYHRYVEDVQLKDLGVKIVYRFSISWP 82
3ahy 8 -----MLPKDFQWNGFA T AAY Q IEGAVDQDGRGPSIWDTFCA-QPG---KIADGSSGVTAACDSYNRTEADIALKSLGAKSYRFSISWS 86
2e3z 6 -----AAKLKPSFVNGYA T AAY Q IEGSPDKDGRGPSIWDTFCK-APG---KIADGSSGDVATDSYNRWEDVQLKSYGVKAYRFSISWS 86
1v02 6 ---RLSPWEI---PRRDWFPSPFLFGAA T SAY Q IEGAWNEDGKGPSTWDHFCNHFPE---WIVDRSNGDVAADSYHMYAEDVRLKEMGMDAYRFSISWP 154
3vif 74 -----DTV-----YTFPDEFKLGAA T ASY Q IEGAWDENGKGNPWDTLTHEHPD---YVVDGATGDIADDSYHLYKEDVQLKELGAQVYRFSISWA 90
1cbg 1 FKPLPISFDDFSDLNRSCFAPGFVFGTA S SAF Q YEGAAFEDGKGPSIWDTFTHKYPE---KIKDRNTGDVAIDYHRYKEDIGIMKDMNLDAYRFSISWP 97
2wbj 25 -----VKKFPEGLWNGVA T ASY Q IEGSPDLADGAGMSIWDTFSHTP-G---NVKNGDTGDVADCHYNRWKEDEIEIEKLGKAYRFSISWP 105
2jfe 3 -----FPAGFGWAAA T AAY Q VEGWADADGKGPSTWDTHQGGE---RVFKNTGDVACGSYTLWEEDLCKIKQLGLTHYRFSISWS 81
5foo 1 -----MLAFPKEFWNGGA T SGP Q SEG--RFAKQHRNLFDYWYEEEPD---LFYDYVGPDTASDAYHQIESDLTLLASLGHNSYRFSISQWT 80
5dt5 25 -----KFAPNFVFGTA T SSY Q IEGAHDEGGRTPSIWDTFCD-TDG---KVFKEHNGDVACDHYHRYFEEDIQHKLQGLVDYRFSIAWP 103
5cg0 2 -----QRRFPDDFLFGTA T ASY Q IEGAWDEDGKGENIWDVYMHNTPE---VIRDLNNGDIAADSYHNYKRDVEMMRELGLDAYRFSISWA 83
4qlk 10 -----NWLGLSRAAFPKRFVFGTV T SAY Q VEGMAASGGKGPSTWDFAH-TPG---NVAGNQGNDVATDQYHRYKEDVNLKMSLNFDAYRFSISWS 97
4ptx 5 -----IIFPEDFIWNGA T SSY Q IEGAFNEDGKGESIWDRTFSH-TPG---KIENGDTGDIADCHYHLYREDIELMKEIGIRSYRFSISWP 84
1vff 1 -----MPLKFPFEMFLFGTA T SSH Q IEGNNR-----WNDDWYEQ-IGK-----LPYRS-GKACNHWEYRDDIQLMTSLGYNAYRFSISWS 74

3wq8 85 RIFPKPTFDVKVDVEKDEEGNIISVDVPESTIKELEKIANMEALEHYRKIYSDWKERKGTFILNLYHW PLPLNIHDP IAVRKLGPDAAPAGWLDE-KTVV 183
3wh5 84 RIVPLGAG-----PINQAGLDHYSRMVDALLGAGLQPFVTL YHW DLPQPLEDR-----LGWGSR-ATAT 141
1gon 96 RIQPTGRG-----PALQKGLDFYRRLADELLAKGIQPVATL YHW DLPQLENA-----GGWPER-ATAE 153
3w53 123 RVVPGGR-----TVNPEGLDFYSRLVDELLENGILPWLTL YHW DLPQALEER-----GGWTNR-ETSY 179
2o9p 92 RIMPAAG-----IINEEGLLFYHLLDEIELAGLIPMLTL YHW DLPQWIEDE-----GGWTOR-ETIQ 148
3cmj 104 RIQPDSSR-----QINQKGLDFYRRLVEGLHKRDLIPMATL YHW DLPQWVEDE-----GGWLSR-ESAS 161
3ahx 84 RIFPKGFG-----EINQKGIQFYRLDIDELIKNDIEPAITL YHW DLPQKLQDI-----GGWANP-QVAD 141
1qox 83 RVLPGQTG-----EVNRAGLDYHRLVDELLANGIEPFCTL YHW DLPQALQDQ-----GGWGSR-ITID 140
3ahy 87 RIIEPGRG-----DAVNQAGIDHYVYFVDLLDAGITPFITL FHW DLPQGLHORY-----GGLLNTEFFL 148
2e3z 87 RIIPKGRS-----DPVNGAGIKHYRTLIEELVKEGITEPFVTL YHW DLPQALDDRY-----GGWLNKEEAIQ 148
1v02 155 RILPKGTLA-----GGINEKRVEYNNKILDLLENGIEPYITL FHW DLPQALVDAY-----GGFLDE-RIIK 215
3vif 91 RVLPEGHD-----NIVNQGDIDYNNLINELLANGIEPMVT M YHW DLPQALQDL-----GGWPNL-VLAK 149
1cbg 98 RVLPGKLS-----GGVNRGINIYNNLINEVLANGMOPYVTL FHW DLPQALEDEY-----RGFLGR-NIVD 158
2wbj 106 RILPEGTG-----RVNQKGLDFYNNIIDLLEKGITPFVTL YHW DLPQALQLK-----GGWANR-EIAD 163
2jfe 82 RILPDGTT-----GFINQKGIQFYNNKILDLLENGIEPYITL YHW DLPQLEDO-----GGWLSR-AIEI 140
5foo 81 RLIDDFEQ-----ATINPDGLAYNVRVIDACLANGIRPVNLHHLF DLPQALQAY-----GGWESK-HVVD 140
5dt5 104 RIFPSK-----GQFNPEGMAFYKTLATRLQEEGKPAVTL YHW DLPQMAHEE-----GGWVNR-DSVD 160
5cg0 84 RILPTGMA-----NEVNPAGIAFYNNIIDLLENGIEPYITL YHW DLPQALQEL-----GGFANL-LISD 142
4qlk 98 RIFPDGE-----GRVNPQEGVAYNNLINYLQKGITPYVNL YHW DLPQALEKKY-----GGWLNK-KMAD 156
4ptx 85 RILPEGK-----GRVNQKGLDFYKRLVDNLLKANIIPMITL YHW DLPQALQDK-----GGWTNR-DTAK 142
1vff 75 RLFPPEEN-----KFNEADAFMKYREIIDLTLRGITPLVTL H H F TSPLEWFMKK-----GGFLRE-ENLK 131

3wq8 184 EFVKFAAFVAYHLDLVDMMSTINEPNVVYNQGINLASGFPFGFL-----SFEAAEKAFNLIQAHIGAYDAIKEYSE-----KSVGVIIYA F 266
3wh5 142 VFAEYADIVVRQLGDRVTHWATLINEPWC SAMLYGGLG--VHAPGHT-----DLKRGLEASHNLLGHGLAVQAMRAAPQF--LQIGIVLNL 224
1gon 154 RFAEYAAIADALGDRVKTWITLINEPWC SAFLYGSG--VHAPGRT-----DPVAALRAAHNLLGHGLAVQALRDLRDLAD--AQCSVTNLN 236
3w53 180 KFLEYAETVHEKLGDRVKHWITFNPCLCSLIGYAAG--EHAPGRQ-----EPQAAALAVHHQHLAHLATARLRELGA--EHIGITLNL 260
2o9p 149 HFKTYASVIMDRFGERINWNTINEPYCASILGYGTG--EHAPGLK-----NWREAFATAHHLMCHGIASNLHKEKGLT--GKIGITLNM 230
3cmj 162 RFAEYTHALVAALGDQIPLVTHNEPMVTVMWAGYHMG--LFAPGLK-----DPTLGGRVAHLLSHGQALQAFRALSPAG--SQMGITLNF 244
3ahx 142 YVYDVANLLFREFGDRVKTWITLINEPWC VAVSYLGYALG--VHAPGIK-----DMKALLAAHNLISHFRAVAVKVFRETVDK--GKIGITLNL 223
1qox 141 AFAEYAEIMFKELGGKIKQWITFNEPWCMAFLSNYLG--VHAPGNK-----DLQLAIDVSHHLLVAHGRAVTLFRELGIS--GEIGIAPNT 222
3ahy 149 DFENYARVMFRA-LPKVRNWTITFNEPLCSAIPGYSG--TFAPGRQ-----STSEPTWVGHNLVAVHGRAVKAYRDFKPFASGGDGIIGIVLNG 233
2e3z 149 WFEYARVLFENFGDRVKNWITFNEPWC I SVMGYGNG--TFAPGHV-----SNTPEWIVSHHLLAHAAVAVKLYRDEPFKEKQG-GGIGITLDS 233
1v02 216 DYTDFAKVCFEFGKTVKNWITFNEPETECSVSYGTG--VLAPGRCSPGVSCAVPTGNSLSEPIYVAHNLRAHAETVDIYNKYHKGAD--GRIGLALNV 311
3vif 150 YSENYARVLKFNFGDRVKLWITFNEPLTMDGYASEI--GMAPSIN-----TPGIGDYLAHTVIHAHARIYHLYDQEFRAEQG-KGVGISLNI 235
1cbg 159 DFRDYAELCFKEFGDRVKHWITLINEPWCVMNAYAG--TFAPGRCSDWLKLNCTGGDSGREPYLAHYQLAHAAARLYTKYQASQN-GIIGITLVS 255
2wbj 164 WFAEYARVLFENFGDRVKNWITLINEPWC VAVSYLGYALG--VHAPGMK-----DIYVAFRAVHNLRAHARAVKVFRETVDK--GKIGITLNL 244
2jfe 141 SFDKYAQCFSTFGDRVKWITLINEANVLSVMSYDLG--MFPPGIP-----HFGTGGYQAHHNLKAKHARSWSYDLSFRKKQK-GMVLSLFA 226
5foo 141 LFVAESKVCQFQFQDRVKWITLINEPWC VVEGYSLMQ--FHPAIVD-----GKKAVQVAYNLALATAKVIQAYRRGFAELS-DGRIGITLNL 225
5dt5 163 WFLDFARVCFEELDGIVDSWITLINEPWCAGFLSYHLG--QHAPGHT-----MNEAVRAVHMLLSHGKAVEMLKGFENS--ATPITLNL 243
5cg0 143 WFEYARVLFENFGDRVKWITLINEPRELCFEGYSGA--TKAPILNA-----TAMGAYLCANLVTAAHAKAYLYDREFRPVQ-GGQGITLVS 228
4qlk 157 LFTFYADFCKTFGNRVKHWITFNEPRIVALLGYDQG--TNPPKRTKCA--AGGNSATEPIYVAHNLPSHAAAVARYRTKYQAQ-GQGVIVLDF 249
4ptx 143 YFAEYARLMPFEFNGLVLDWITLINEPWC VAFEGHAFG--NHAPGTD-----FKTALQVAHLLSHGMVAVDIFREEDL--PGEIGITLNL 224
1vff 132 HWKYEIEKVAEL-LEKVKLVATFNEPMVYVMMGYLTA--YWPPFIR-----SPFKAFKVAANLLKAHAIAVELLHGK-----FKVGIVKNI 209

Figura 95. Alinhamento de sequências de diferentes beta-glicosidases GH1 com destaque para os resíduos centrais por *closeness*. Os resíduos com valor de centralidade maior do que 2,5 desvios padrão da média estão destacados com fundo cinza escuro e maior do que 2 desvios estão destacados com fundo cinza. Regiões de alfa hélices estão em vermelho e fitas beta em azul.

5.4 Conclusões

As redes de estrutura de Sfbgli construídas segundo o critério de contato entre resíduos baseado em carbonos-alfa a 8,5 Å ou quaisquer átomos a 5 Å são muito semelhantes. Adicionalmente observa-se que todas as redes geradas com contatos em distâncias de corte maiores do que 4 Å tem características semelhantes. De fato, a partir deste limite de distâncias de corte já se observam as características de “Mundo Pequeno” da rede estrutural de Sfbgli, com C_p (0,48) muito maior que aqueles de redes randômicas (0,02) e regulares e L_p (4,8) semelhante àqueles de uma rede randômica (2,6), mas muito menor que uma rede regular (Watts e Strogatz, 1998).

A decomposição da rede estrutural de Sfbgli nas subredes de interações de curto e longo alcance mostra que a principal componente da agregação (C_p alto) está nos contatos entre resíduos próximos na estrutura primária, enquanto que o caminho mínimo médio curto (L_p pequeno) ocorre principalmente graças às conexões de longo alcance. As distribuições de contatos destas subredes não se ajustam em uma distribuição de Poisson ou Lei de Potência, embora na subrede de longo alcance seja mais provável encontrar resíduos muito conectados do que na subrede de curto alcance. Esta característica aproxima a subrede de longo alcance de uma rede “Sem Escala”, especialmente levando em consideração que a superfície finita dos resíduos de aminoácidos limita o número de contatos possíveis.

A análise da rede estrutural de Sfbgli segundo diferentes critérios de centralidade mostra que de forma geral o parâmetro de centralidade por grau é semelhante entre os resíduos da rede estrutural, seguindo uma distribuição normal. Isto indica que não existem diferenças significativas entre os resíduos, somente oscilações ao acaso e consequentemente os resíduos

centrais por grau são regularmente espalhados ao longo de toda a rede estrutural. A centralidade dos resíduos por *closeness* também segue uma distribuição normal. Porém, sua localização mostra tendência de concentração em pontos específicos da rede estrutural de Sfβgli, dos quais tendem a partir muitas conexões de longo alcance que os interconectam. Já a distribuição da centralidade por *betweeness* na rede estrutural de Sfβgli desvia significativamente de uma curva normal, e existe uma população de resíduos com centralidade maior do que 3 desvios padrão da média. Esse mesmo tipo de distribuição também ocorre para a centralidade $C_{\Delta Lp}$.

O mapeamento dos resíduos centrais por *closeness* na estrutura secundária e terciária de Sfβgli mostra que eles normalmente ocorrem na região final de fitas-beta (3) e alças (10) posterior a estas fitas beta. Apenas um resíduo central por *closeness* ocorre em alfa-hélice. O mesmo padrão é observado para os resíduos centrais por *betweeness*, ocorrendo mais frequentemente em fitas beta (2) e alças (8) do que em alfa-hélices (2). Os resíduos centrais por $C_{\Delta Lp}$ acompanham a distribuição dos resíduos centrais por *closeness*, ficando concentrados em fitas-beta (6) e alças (4) na metade C-terminal de Sfβgli. Apenas os resíduos com alto grau estão mais concentrados em alfa-hélices (5).

A correlação de centralidade com propriedades mostra que os resíduos centrais por *closeness* tendem a se concentrar no sítio ativo e pontos menos flexíveis da estrutura de Sfβgli.

Por fim, a construção de as redes para diversas beta-glicosidases GH1 mostra que os parâmetros C_p , L_p e o número médio de conexões são bastante próximos entre todas elas, assim como a distribuição dos resíduos centrais por *closeness*. O alinhamento de sequências destas

beta-glicosidases GH1 revela razoável conservação dos resíduos que ocupam as posições centrais por *betweenness* e *closeness*.

CONCLUSÃO GERAL

Diversos trabalhos têm proposto que as Redes de Estrutura Proteica são ferramentas promissoras para caracterização de proteínas, auxiliando a compreender o papel dos resíduos de aminoácidos nas propriedades e função de proteínas (Vendruscolo *et al.*, 2001; Greene e Higman, 2003; Del Sol e O'meara, 2005; Del Sol *et al.*, 2006; Di Paola e Giuliani, 2015; Fokas *et al.*, 2016). No entanto, na literatura, não haviam trabalhos desenhados para verificar esta hipótese. Neste sentido, esta tese elucida e ressalta a importância dos resíduos centrais na manutenção de propriedades e função de proteínas.

Inicialmente, como descrito no capítulo 1, foi feito um desenho experimental que permitisse avaliar o efeito de perturbações em resíduos centrais e em resíduos não centrais, de forma a avaliar suas diferenças em cada grupo. Para isso, a beta-glicosidase GH1 de *Spodoptera frugiperda* (Sfβgli) foi utilizada como modelo experimental, e mutações foram desenhadas para modificar cada um de seus resíduos centrais por alanina (que possui cadeia lateral pequena, fazendo poucas interações com a vizinhança). O mesmo foi feito para um grupo de resíduos não centrais. Similarmente mutações foram introduzidas na vizinhança do resíduo central F251. Após isso, foi feita a produção e purificação destas proteínas mutantes recombinantes, bem como análises estruturais que demonstraram que todas se apresentavam enoveladas.

Além disso, conforme o capítulo 2, a estabilidade térmica dessas proteínas foi determinada, primeiramente através da determinação das suas T_m , e posteriormente através da caracterização da velocidade de desnaturação. Em ambas abordagens, observou-se que as mutações diretas em resíduos centrais diminuíram a estabilidade térmica da proteína, enquanto que as mutações em resíduos não centrais praticamente não tiveram efeitos. Além disso,

mutações na vizinhança do resíduo central F251 também afetaram a termoestabilidade da Sfβgli. Mutações em resíduos cuja cadeia lateral estava diretamente apontada para o resíduo F251 tiveram grandes efeitos na estabilidade térmica da enzima, diminuindo a T_m em 15°C no caso da mutação V336F, ou até mesmo provocando a inexistência de transição na desnaturação, como no caso da mutação A264F. Por outro lado, mutações em resíduos com cadeia lateral apontada em direções opostas ao resíduo F251 praticamente não tiveram efeito na termoestabilidade da proteína.

Além desta análise, possíveis mudanças na cinética de catálise pela enzima devido às mutações foram avaliadas no Capítulo 3. Para isso, Sfβgli com mutações que perturbassem o resíduo central F251 foram avaliadas. De uma forma geral, observou-se perda de atividade enzimática provocada por estas mutações, principalmente devido à redução do k_{cat} . Curiosamente, também se observou que o efeito de redução do k_{cat} parece relacionado à perda de termoestabilidade, indicando que as mutações na vizinhança do resíduo central foram propagadas pela estrutura proteica, diminuindo a sua estabilidade térmica, e atingindo o sítio ativo, diminuindo o k_{cat} da Sfβgli.

No capítulo 4 foi feita uma investigação sobre a estrutura quaternária da Sfβgli e sua atividade enzimática, incluindo a importância dos resíduos centrais nesta correlação. Para isso, inicialmente, o arranjo oligomérico da Sfβgli em solução foi avaliado por SAXS, revelando que a proteína se apresenta majoritariamente como dímero em tampão citrato-fosfato pH 6,0 100 mM (CF100), com uma pequena fração da população como tetrâmero. Por outro lado, em tampão fosfato de potássio pH 6,0 10 mM (F10), observou-se que Sfβgli apresenta-se como um grande arranjo oligomérico com cerca de 12 unidades monoméricas. Além disso, observou-se que a

Sf β gli em tampão F10 sofre um processo lento de ativação, tendo como resultado uma amostra com atividade semelhante à enzima em CF100. Levantou-se a hipótese que esta ativação estaria correlacionada a uma mudança do estado oligomérico da Sf β gli em F10 (grandes oligômeros) para um estado semelhante ao da Sf β gli em CF100 (dímeros). Por fim, ainda no capítulo 4, os arranjos oligoméricos da Sf β gli foram analisados utilizando a abordagem de redes, sendo observado que o aumento do grau de oligomerização da Sf β gli faz com que a região com alta centralidade por *closeness*, que no monômero localiza-se no sítio ativo, passe a ser localizado na interface dimérica ou na interface tetramérica (no centro, entre dois dímeros).

Já no Capítulo 5, a centralidade dos resíduos de diversas beta-glicosidases GH1 de Bacteria, Archea e Eucariotos foi determinada e sua conservação analisada através de alinhamentos e comparações estruturais. De uma forma geral, observou-se grande conservação dos resíduos, especialmente ao avaliar a centralidade por *closeness*, sendo que tanto o tipo de resíduo parece conservado, quanto sua posição nas sequências das beta-glicosidases. Adicionalmente, notou-se que resíduos centrais por *closeness*, *betweenness* e $C_{\Delta Lp}$ tendem a ocorrer no final (C-terminal) de fitas beta ou alças que a elas se seguem, estando assim concentrados preferencialmente no topo do barril onde localiza-se o sítio ativo.

Combinando todas estas observação, temos que de uma forma geral, os resíduos centrais por $C_{\Delta Lp}$ parecem desempenhar papéis fundamentais para a estabilidade térmica da Sf β gli, indicando que perturbações nestes centros podem mudar o fluxo de energia térmica ao longo da estrutura proteica. Além disso, observa-se que estas mesmas mutações diminuíram a atividade catalítica, revelando que as perturbações são propagadas e atingem o sítio ativo. Finalmente,

arranjos oligoméricos podem influenciar a atividade enzimática da Sfβgli e modificar a topologia da sua rede estrutural, mudanças que poderiam, hipoteticamente, estar relacionadas.

Por fim, conclui-se por este trabalho que, de fato, a análise de proteínas através de redes é uma ferramenta que pode revelar o papel dos seus resíduos na determinação de suas propriedades e função. Esta conclusão é coerente com o fenômeno observado por Albert e colaboradores (2000), onde deleções aleatórias dos nós de redes “Mundo Pequeno e Sem Escala” geram pouco efeito na rede como um todo. No entanto, deleções direcionadas aos nós centrais causam efeitos claramente deletérios ao funcionamento da rede. Deste modo, analogamente a topologia das Redes de Estrutura Proteica pode ser determinante para robustez das proteínas frente a mutações ao acaso, mas colapso funcional em perturbações nos resíduos centrais.

REFERÊNCIAS BIBLIOGRÁFICAS

AFTABUDDIN, M.; KUNDU, S. Hydrophobic, hydrophilic, and charged amino acid networks within protein. **Biophysical Journal**, v. 93, n. 1, p. 225-231, Jul 2007. ISSN 0006-3495.

ALBERT, R.; JEONG, H.; BARABASI, A. L. Error and attack tolerance of complex networks. **Nature**, v. 406, n. 6794, p. 378-382, Jul 2000. ISSN 0028-0836.

AMITAI, G. et al. Network analysis of protein structures identifies functional residues. **Journal of Molecular Biology**, v. 344, n. 4, p. 1135-1146, Dec 2004. ISSN 0022-2836.

ANTWERPEN, M. et al. Whole-genome sequencing of a pandoravirus isolated from keratitis-inducing acanthamoeba. **Genome announcements**, v. 3, n. 2, p. e00136-15, 2015. ISSN 2169-8287.

ARNOLD, F. H. et al. How enzymes adapt: lessons from directed evolution. **Trends in biochemical sciences**, v. 26, n. 2, p. 100-106, 2001. ISSN 0968-0004.

ATILGAN, A. R.; AKAN, P.; BAYSAL, C. Small-world communication of residues and significance for protein dynamics. **Biophysical Journal**, v. 86, n. 1, p. 85-91, Jan 2004. ISSN 0006-3495.

BACHMAIER, C.; BRANDES, U.; SCHREIBER, F. **Biological networks**. 2014. ISBN 1584884126.

BAGLER, G.; SINHA, S. Network properties of protein structures. **Physica a-Statistical Mechanics and Its Applications**, v. 346, n. 1-2, p. 27-33, Feb 1 2005. ISSN 0378-4371.

BANERJEE, R. et al. CRYSTAL-STRUCTURE OF PEANUT LECTIN, A PROTEIN WITH AN UNUSUAL QUATERNARY STRUCTURE. **Proceedings of the National Academy of Sciences of the United States of America**, v. 91, n. 1, p. 227-231, Jan 1994. ISSN 0027-8424.

BARABASI, A.-L. **LINKED - A NOVA CIENCIA DOS NETWORKS**. 2009. ISBN 8528906124.

BERG, J. M.; TYMOCZKO, J. L.; STRYER, L. **Biochemistry**. 7th edition. W. H. Freeman, 2012.

BORNSCHEUER, U. T. et al. Engineering the third wave of biocatalysis. **Nature**, v. 485, n. 7397, p. 185-194, May 2012. ISSN 0028-0836.

BRINDA; VISHVESHWARA, S.; VISHVESHWARA, S. Random network behaviour of protein structures. **Molecular Biosystems**, v. 6, n. 2, p. 391-398, 2010 2010. ISSN 1742-206X.

BRINDA, K. V.; VISHVESHWARA, S. A network representation of protein structures: Implications for protein stability. **Biophysical Journal**, v. 89, n. 6, p. 4159-4170, Dec 2005. ISSN 0006-3495.

BROGNARO, H. et al. Biochemical Characterization and Low-Resolution SAXS Molecular Envelope of GH1 β -Glycosidase from *Saccharophagus degradans*. **Molecular biotechnology**, v. 58, n. 12, p. 777-788, 2016. ISSN 1073-6085.

BÖDE, C. et al. Network analysis of protein dynamics. **Febs Letters**, v. 581, n. 15, p. 2776-2782, Jun 19 2007. ISSN 0014-5793.

CAIRNS, J. R. K.; ESEN, A. β -Glucosidases. **Cellular and Molecular Life Sciences**, v. 67, n. 20, p. 3389-3405, 2010. ISSN 1420-682X.

CAPRA, J. A.; SINGH, M. Predicting functionally important residues from sequence conservation. **Bioinformatics**, v. 23, n. 15, p. 1875-1882, Aug 2007. ISSN 1367-4803.

CHAIRES, J. B. Calorimetry and thermodynamics in drug design. In: (Ed.). **Annual Review of Biophysics**: Annual Reviews, v.37, 2008. p.135-151. (Annual Review of Biophysics). ISBN 978-0-8243-1837-6.

CHENG, S.; FU, H.-L.; CUI, D.-X. Characteristics analyses and comparisons of the protein structure networks constructed by different methods. **Interdisciplinary Sciences: Computational Life Sciences**, v. 8, n. 1, p. 65-74, 2016. ISSN 1913-2751.

COLUSSI, F. et al. Oligomeric state and structural stability of two hyperthermophilic β -glucosidases from *Thermotoga petrophila*. **Amino acids**, v. 47, n. 5, p. 937-948, 2015. ISSN 0939-4451.

CZIZEK, M. et al. The mechanism of substrate (aglycone) specificity in beta-glucosidases is revealed by crystal structures of mutant maize beta-glucosidase-DIMBOA, -DIMBOAGlc, and -dhurrin complexes. **Proceedings of the National Academy of Sciences of the United States of America**, v. 97, n. 25, p. 13555-13560, Dec 2000. ISSN 0027-8424.

DAVIES, G. J.; WILSON, K. S.; HENRISSAT, B. Nomenclature for sugar-binding subsites in glycosyl hydrolases. **Biochemical Journal**, v. 321, p. 557-559, Jan 1997. ISSN 0264-6021.

DEL SOL, A. et al. Residue centrality, functionally important residues, and active site shape: Analysis of enzyme and non-enzyme families. **Protein Science**, v. 15, n. 9, p. 2120-2128, 2006. ISSN 1469-896X.

DEL SOL, A.; O'MEARA, P. Small-world network approach to identify key residues in protein-protein interaction. **Proteins-Structure Function and Bioinformatics**, v. 58, n. 3, p. 672-682, Feb 15 2005. ISSN 0887-3585.

DEL_SOL et al. Residues crucial for maintaining short paths in network communication mediate signaling in proteins. **Mol Syst Biol**, v. 2, p. 2006 0019, 2006. ISSN 1744-4292.

DI PAOLA, L.; GIULIANI, A. Protein contact network topology: a natural language for allostery. **Current opinion in structural biology**, v. 31, p. 43-48, 2015. ISSN 0959-440X.

DUNITZ, J. D. Win Some, Lose Some - Enthalpy-Entropy Compensation in Weak Intermolecular Interactions. **Chemistry and Biology**, v. 2, n. 11, p. 709-712, Nov 1995. ISSN 1074-5521.

EFTINK, M. R.; GHIRON, C. A. Fluorescence Quenching Studies With Proteins. **Analytical Biochemistry**, v. 114, n. 2, p. 199-227, 1981. ISSN 0003-2697.

ESCAMILLA-ALVARADO, C. et al. An overview of the enzyme potential in bioenergy-producing biorefineries. **Journal of Chemical Technology and Biotechnology**, v. 92, n. 5, p. 906-924, May 2017. ISSN 0268-2575.

ESEN, A. β -Glucosidases. In: (Ed.): ACS Publications, 1993. ISBN 1947-5918.

FEIGIN, L. A.; SVERGUN, D. I. Structure Analysis by Small-Angle X-Ray and Neutron Scattering. Boston, MA, 1987. ISSN 9781475766240 1475766246.

FERSHT, A. **Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding**. W. H. Freeman, 1999. ISBN 9780716732686.

FOKAS, A. S. et al. Residue Geometry Networks: A Rigidity-Based Approach to the Amino Acid Network and Evolutionary Rate Analysis. **Scientific Reports**, v. 6, 2016.

FRUTUOSO, M. A.; MARANA, S. R. A Single Amino Acid Residue Determines the Ratio of Hydrolysis to Transglycosylation Catalyzed by beta-Glucosidases. **Protein and Peptide Letters**, v. 20, n. 1, p. 102-106, Jan 2013. ISSN 0929-8665.

GLATTER, O.; KRAKTY, O. **Small Angle X-Ray Scattering**. London: Academic Press, 1982.

GREENE, L. H. Protein structure networks. **Briefings in Functional Genomics**, v. 11, n. 6, p. 469-478, Nov 2012. ISSN 2041-2649.

GREENE, L. H.; HIGMAN, V. A. Uncovering network systems within protein structures. **Journal of Molecular Biology**, v. 334, n. 4, p. 781-791, Dec 5 2003. ISSN 0022-2836.

GUINIER, A.; FOURNET, G. **Small-angle scattering of X-rays**. Wiley, 1955.

IKEGAMI, C. **Caracterização funcional dos resíduos centrais da rede estrutural da b-glicosidase SfbgI de *Spodoptera frugiperda***. Instituto de Química: Universidade de São Paulo, 2013. 103 Disponível em: < Biblioteca Digital de Teses e Dissertações da USP. <http://www.teses.usp.br/teses/disponiveis/46/46131/tde-21052013-134648/pt-br.php> >.

ISORNA, P. et al. Crystal structures of *Paenibacillus polymyxa* beta-glucosidase B complexes reveal the molecular basis of substrate specificity and give new insights into the catalytic machinery of family I Glycosidases. **Journal of Molecular Biology**, v. 371, n. 5, p. 1204-1218, Aug 2007. ISSN 0022-2836.

JANIN, J.; BAHADUR, R. P.; CHAKRABARTI, P. Protein-protein interaction and quaternary structure. **Quarterly Reviews of Biophysics**, v. 41, n. 2, p. 133-180, May 2008. ISSN 0033-5835.

KANNAN, N.; VISHVESHWARA, S. Identification of side-chain clusters in protein structures by a graph spectral method. **Journal of molecular biology**, v. 292, n. 2, p. 441-464, 1999. ISSN 0022-2836.

KELLY, S. M.; JESS, T. J.; PRICE, N. C. How to study proteins by circular dichroism. **Biochimica Et Biophysica Acta-Proteins and Proteomics**, v. 1751, n. 2, p. 119-139, Aug 10 2005. ISSN 1570-9639.

KHANIN, R.; WIT, E. How scale-free are biological networks. **Journal of Computational Biology**, v. 13, n. 3, p. 810-818, Apr 2006. ISSN 1066-5277.

KIM, S.; KIM, I. Oligomeric structure of β -glucosidases. **J. Photosci**, v. 11, p. 121-127, 2004.

KIM, Y.-W. et al. Formation of fibrillar multimers of oat β -glucosidase isoenzymes is mediated by the As-Glu1 monomer. **Journal of molecular biology**, v. 303, n. 5, p. 831-842, 2000. ISSN 0022-2836.

KIRK, O.; BORCHERT, T. V.; FUGLSANG, C. C. Industrial enzyme applications. **Current Opinion in Biotechnology**, v. 13, n. 4, p. 345-351, Aug 2002. ISSN 0958-1669.

KRISSINEL, E.; HENRICK, K. Inference of macromolecular assemblies from crystalline state. **Journal of Molecular Biology**, v. 372, n. 3, p. 774-797, Sep 2007. ISSN 0022-2836.

KRUJA, E. et al. A Short Note on the History of Graph Drawing. In: MUTZEL, P.; JÜNGER, M., *et al* (Ed.). **Graph Drawing: 9th International Symposium, GD 2001 Vienna, Austria, September 23–26, 2001 Revised Papers**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2002. p.272-286. ISBN 978-3-540-45848-7.

LADERO, M.; SANTOS, A.; GARCIA-OCHOA, F. Kinetic modelling of the thermal inactivation of an industrial beta-galactosidase from *Kluyveromyces fragilis*. **Enzyme and Microbial Technology**, v. 38, n. 1-2, p. 1-9, Jan 3 2006. ISSN 0141-0229.

LAEMMLI, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. **Nature**, v. 227, n. 5259, p. 680-5, Aug 15 1970. ISSN 0028-0836.

LEHMANN, M. et al. The consensus concept for thermostability engineering of proteins. **Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology**, v. 1543, n. 2, p. 408-415, 2000. ISSN 0167-4838.

LEITNER, D. M. Energy flow in proteins. **Annu. Rev. Phys. Chem.**, v. 59, p. 233-259, 2008. ISSN 0066-426X.

LEVINE, I. **Physical Chemistry**. 6ed. New York: I, 2009. 989.

LEVY, E. D. PiQSi: Protein quaternary structure investigation. **Structure**, v. 15, n. 11, p. 1364-1367, Nov 2007. ISSN 0969-2126.

LI, Y. et al. Predicting disease-associated substitution of a single amino acid by analyzing residue interactions. **BMC bioinformatics**, v. 12, n. 1, p. 14, 2011. ISSN 1471-2105.

LOCKLESS, S. W.; RANGANATHAN, R. Evolutionarily conserved pathways of energetic connectivity in protein families. **Science**, v. 286, n. 5438, p. 295-299, 1999. ISSN 0036-8075.

LOMBARD, V. et al. The carbohydrate-active enzymes database (CAZy) in 2013. **Nucleic acids research**, v. 42, n. D1, p. D490-D495, 2014. ISSN 0305-1048.

LOUIS-JEUNE, C.; ANDRADE-NAVARRO, M. A.; PEREZ-IRATXETA, C. Prediction of protein secondary structure from circular dichroism using theoretically derived spectra. **Proteins: Structure, Function, and Bioinformatics**, v. 80, n. 2, p. 374-381, 2012. ISSN 1097-0134.

LUAN, C.-H. et al. Ligand Screening Using Fluorescence Thermal Shift Analysis (FTS). In: ANDERSON, W. F. (Ed.). **Structural Genomics and Drug Discovery**. Northwestern University Feinberg, School of Medicine, Chicago, USA: Springer Protocols, 2014. cap. 20,

MALLIK, S.; KUNDU, S. Modular organization of residue-level contacts shape the selection pressure on individual amino acid sites of ribosomal proteins. **Genome Biology and Evolution**, v. 9, n. 4, p. 916-931, 2017.

MARANA, S. R. Molecular basis of substrate specificity in family 1 glycoside hydrolases. **Iubmb Life**, v. 58, n. 2, p. 63-73, Feb 2006. ISSN 1521-6543.

MARANA, S. R. et al. Amino acid residues involved in substrate binding and catalysis in an insect digestive beta-glycosidase. **Biochimica Et Biophysica Acta-Protein Structure and Molecular Enzymology**, v. 1545, n. 1-2, p. 41-52, Feb 9 2001. ISSN 0167-4838.

MARANA, S. R.; TERRA, W. R.; FERREIRA, C. Purification and properties of a beta-glycosidase purified from midgut cells of *Spodoptera frugiperda* (Lepidoptera) larvae. **Insect Biochemistry and Molecular Biology**, v. 30, n. 12, p. 1139-1146, Dec 2000. ISSN 0965-1748.

MARANA, S. R. et al. The role of residues R97 and Y331 in modulating the pH optimum of an insect beta-glycosidase of family 1. **European Journal of Biochemistry**, v. 270, n. 24, p. 4866-4875, Dec 2003. ISSN 0014-2956.

MENDONÇA, L. M. F.; MARANA, S. R. The role in the substrate specificity and catalysis of residues forming the substrate aglycone-binding site of a beta-glycosidase. **Febs Journal**, v. 275, n. 10, p. 2536-2547, May 2008. ISSN 1742-464X.

MENDONÇA, L. M. F.; MARANA, S. R. Single mutations outside the active site affect the substrate specificity in a beta-glycosidase. **Biochimica Et Biophysica Acta-Proteins and Proteomics**, v. 1814, n. 12, p. 1616-1623, Dec 2011. ISSN 1570-9639.

MERTENS, H. D. T.; SVERGUN, D. I. Structural characterization of proteins and complexes using small-angle X-ray solution scattering. **Journal of Structural Biology**, v. 172, n. 1, p. 128-141, Oct 2010. ISSN 1047-8477.

MIYAZAKI, K. et al. Directed evolution study of temperature adaptation in a psychrophilic enzyme. **Journal of molecular biology**, v. 297, n. 4, p. 1015-1026, 2000. ISSN 0022-2836.

MORITA, H.; TAKANO, M. Residue network in protein native structure belongs to the universality class of a three-dimensional critical percolation cluster. **Physical Review E**, v. 79, n. 2, p. 020901, 2009.

NIESEN, F. H.; BERGLUND, H.; VEDADI, M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. **Nature Protocols**, v. 2, n. 9, p. 2212-2221, 2007. ISSN 1754-2189.

NIKNAM, N. et al. PDB2Graph: A toolbox for identifying critical amino acids map in proteins based on graph theory. **Computers in biology and medicine**, v. 72, p. 151-159, 2016. ISSN 0010-4825.

OLIVEIRA, C. L. Investigating Macromolecular Complexes in Solution by Small Angle X-Ray Scattering. In: CHANDRASEKARAN, D. A. (Ed.). **Current Trends in X-Ray Crystallography**: InTech, 2011.

OLIVEIRA, C. L. P. **Estudo de Espalhamento de Raios X a Baixos Ângulos por Sistemas Biológicos: Teoria e Aplicações**. 2005. IFGW, UNICAMP

PANTOLIANO, M. W. et al. High-density miniaturized thermal shift assays as a general strategy for drug discovery. **Journal of biomolecular screening**, v. 6, n. 6, p. 429-440, 2001. ISSN 1087-0571.

Protein Data Bank. <http://www.rcsb.org/pdb/home/home.do>.

ROZENFELD, J. H. K. **Estudo das bases moleculares da especificidade pelo substrato de uma beta-glicosidase (AF052729) de Spodoptera frugiperda**. 2006. Universidade de São Paulo. Disponível na Biblioteca Digital de Teses e Dissertações da USP. <http://www.teses.usp.br/teses/disponiveis/46/46131/tde-19092006-163823/pt-br.php>

RYE, C. S.; WITHERS, S. G. Glycosidase mechanisms. **Current opinion in chemical biology**, v. 4, n. 5, p. 573-580, 2000. ISSN 1367-5931.

SHANK, E. A. et al. The folding cooperativity of a protein is controlled by its chain topology. **Nature**, v. 465, n. 7298, p. 637-U134, Jun 3 2010. ISSN 0028-0836.

SHOICHET, B. K. et al. A relationship between protein stability and protein function. **Proceedings of the National Academy of Sciences**, v. 92, n. 2, p. 452-456, 1995. ISSN 0027-8424.

SINGH, G.; VERMA, A.; KUMAR, V. Catalytic properties, functional attributes and industrial applications of β -glucosidases. **3 Biotech**, v. 6, n. 1, p. 3, 2016. ISSN 2190-572X.

SINGHANIA, R. R. et al. Industrial Enzymes: b-Glucosidases. **Current Developments in Biotechnology and Bioengineering: Production, Isolation and Purification of Industrial Products**, p. 103, 2016. ISSN 0444636730.

SOMERO, G. N. Proteins and temperature. **Annual review of physiology**, v. 57, n. 1, p. 43-68, 1995. ISSN 0066-4278.

SOUZA, V. P. et al. Protein thermal denaturation is modulated by central residues in the protein structure network. **Febs Journal**, v. 283, n. 6, p. 1124-1138, MAR 2016 2016. ISSN 1742-464X;1742-4658.

SUEMORI, A. Conserved and non-conserved residues and their role in the structure and function of p-hydroxybenzoate hydroxylase. **Protein Engineering Design & Selection**, v. 26, n. 7, p. 479-488, Jul 2013. ISSN 1741-0126.

SVERGUN, D. et al. Protein hydration in solution: experimental observation by x-ray and neutron scattering. **Proceedings of the National Academy of Sciences**, v. 95, n. 5, p. 2267-2272, 1998. ISSN 0027-8424.

SVERGUN, D. I. Small-angle scattering studies of macromolecular solutions. **Journal of Applied Crystallography**, v. 40, p. S10-S17, Apr 2007. ISSN 0021-8898.

SZALAY, K. Z.; CSERMELY, P. Perturbation centrality and Turbine: a novel centrality measure obtained using a versatile network dynamics tool. **PloS one**, v. 8, n. 10, p. e78059, 2013. ISSN 1932-6203.

SÜEL, G. M. et al. Evolutionarily conserved networks of residues mediate allosteric communication in proteins. **Nature Structural & Molecular Biology**, v. 10, n. 1, p. 59-69, 2003.

TAMAKI, F. K. et al. Using the Amino Acid Network to Modulate the Hydrolytic Activity of beta-Glycosidases. **Plos One**, v. 11, n. 12, DEC 9 2016 2016. ISSN 1932-6203.

TRIA, G. et al. Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. **IUCrJ**, v. 2, n. 2, p. 207-217, 2015. ISSN 2052-2525.

TSE, A.; VERKHIVKER, G. Small-world networks of residue interactions in the Abl kinase complexes with cancer drugs: topology of allosteric communication pathways can determine drug resistance effects. **Molecular BioSystems**, v. 11, n. 7, p. 2082-2095, 2015.

VASSILENKO, K. S.; UVERSKY, V. N. Native-like secondary structure of molten globules. **Biochimica Et Biophysica Acta-Protein Structure and Molecular Enzymology**, v. 1594, n. 1, p. 168-177, Jan 2002. ISSN 0167-4838.

VENDRUSCOLO et al. Three key residues form a critical contact network in a protein folding transition state. **Nature**, v. 409, n. 6820, p. 641-645, Feb 2001. ISSN 0028-0836.

VENDRUSCOLO, M. et al. Small-world view of the amino acids that play a key role in protein folding. **Physical Review E**, v. 65, n. 6, Jun 2002. ISSN 1539-3755.

VENTURA, S.; VILLAVERDE, A. Protein quality in bacterial inclusion bodies. **Trends in Biotechnology**, v. 24, n. 4, p. 179-185, Apr 2006. ISSN 0167-7799.

VERDOUCQ, L. et al. Structural determinants of substrate specificity in family 1 beta-glucosidases - Novel insights from the crystal structure of sorghum dhurrinase-1, a plant beta-glucosidase with strict specificity, in complex with its natural substrate. **Journal of Biological Chemistry**, v. 279, n. 30, p. 31796-31803, Jul 2004. ISSN 0021-9258.

VIJAYABASKAR, M.; VISHVESHWARA, S. Interaction energy based protein structure networks. **Biophysical journal**, v. 99, n. 11, p. 3704-3715, 2010. ISSN 0006-3495.

VILLAVERDE, A.; CARRIO, M. M. Protein aggregation in recombinant bacteria: biological role of inclusion bodies. **Biotechnology Letters**, v. 25, n. 17, p. 1385-1395, Sep 2003. ISSN 0141-5492.

WARSHEL, A.; FLORIÁN, J. Computer simulations of enzyme catalysis: Finding out what has been optimized by evolution. **Proceedings of the National Academy of Sciences**, v. 95, n. 11, p. 5950-5955, 1998.

WARSHEL, A.; SCHWEINS, T.; FOTHERGILL, M. Linear Free-Energy Relationships in Enzymes-Theoretical-Analysis of the Reaction of Tyrosyl-Transfer-RNA Synthetase. **Journal of the American Chemical Society**, v. 116, n. 19, p. 8437-8442, 1994.

WATTS, D. J.; STROGATZ, S. H. Collective dynamics of 'small-world' networks. **Nature**, v. 393, n. 6684, p. 440-442, Jun 4 1998. ISSN 0028-0836.

WIERENGA, R. The TIM-barrel fold: a versatile framework for efficient enzymes. **FEBS letters**, v. 492, n. 3, p. 193-198, 2001. ISSN 1873-3468.

WUTHRICH, K. PROTEIN-STRUCTURE DETERMINATION IN SOLUTION BY NMR-SPECTROSCOPY. **Journal of Biological Chemistry**, v. 265, n. 36, p. 22059-22062, Dec 1990. ISSN 0021-9258.

YU, H.; HUANG, H. Engineering proteins for thermostability through rigidifying flexible sites. **Biotechnology advances**, v. 32, n. 2, p. 308-315, 2014. ISSN 0734-9750.

ZANPHORLIN, L. M. et al. Oligomerization as a strategy for cold adaptation: structure and dynamics of the GH1 β -glucosidase from *Exiguobacterium antarcticum* B7. **Scientific reports**, v. 6, 2016.

ZUIDERWEG, E. R. P. Mapping protein-protein interactions in solution by NMR Spectroscopy. **Biochemistry**, v. 41, n. 1, p. 1-7, Jan 2002. ISSN 0006-2960.

ANEXOS

SÚMULA CURRICULAR

1. DADOS PESSOAIS

Nome: Valquiria Pianheri Souza

Local e data de nascimento: São Paulo, 10/01/1989.

2. EDUCAÇÃO

EE. Professora Maria Aparecida de Castro Masiero, São Paulo, 2004-2006. Ensino Médio.

ETEC. Getúlio Vargas, São Paulo, 2005-2006. Curso técnico profissionalizante em Química.

Universidade de São Paulo, São Paulo, 2007-2010. Graduação em Química (bacharelado).

Universidade Virtual do Estado de São Paulo, São Paulo, 2015-atual. Graduação em Licenciatura em Ciências Naturais e Química.

3. FORMAÇÃO COMPLEMENTAR

Bolsista PAE no 2º semestre de 2015, na disciplina de Bioquímica Experimental, coordenada pelos prof. Dr. Carlos Hotta e prof. Dr. Ronaldo Quaggio.

Bolsista PAE no 1º semestre de 2015, na disciplina de Biologia Molecular, coordenada pelo prof. Dr. Eduardo Rego Moraes dos Reis.

Bolsista PAE no 2º semestre de 2014, na disciplina de Biologia I, coordenada pelos prof. Dr. Shaker Chuck Farah, prof. Dra. Alícia Kowaltowski e prof. Dr. Sérgio Verjovski.

4. OCUPAÇÃO

Cursando a pós-graduação (Doutorado Direto) em Bioquímica pelo Instituto de Química. Bolsista de Doutorado Direto, FAPESP, 2011-2016.

5. PUBLICAÇÕES

5.1 Artigos Completos

Souza, V.P.; Ikegami, C.M.; Arantes, G.M.; Marana, S.R.; Protein thermal denaturation is modulated by central residues in the protein structure network; FEBS Journal, 283, 6, 2016, p.1124-1138.

Tamaki, F.K.; Souza, D.P.; Souza, V.P; Ikegami, C.M.; Farah, C.S.; Marana, S.R.; Using the Amino Acid Network to Modulate the Hydrolytic Activity of beta-Glycosidases; Plos One, 11, 12, 2016, e0167978.

Tamaki, F.K.; Souza, D.P.; Souza, V.P; Ikegami, C.M.; Farah, C.S.; Marana, S.R.; Depósito de Estrutura no Protein Data Bank (PDB) – Estrutura da Beta-glicosidase de *Spodoptera frugiperda* 5CG0, 2015.

5.2 Artigos em preparação

Souza, V. P.; Ikegami, C. M.; Arantes, G. M.; Marana, S. R.; A central residue in the protein structure network propagates mutational effects into the active site in a b-glucosidase GH1 (em finalização).

Souza, V. P.; Bicev, R.; Oliveira, C. P.; Marana, S. R.; The importance of the oligomerization in the activity of a beta-glucosidase GH1 (em fase inicial de escrita).

Florindo, R. N.; Souza, V. P.; Mutti, H. S.; Camilo, C.; Margarido, L. R. M.; Marana, S. R.; Polikarpov, I.; Nascimento, A. S.; Structural Insights into β -Glucosidase Transglycosylation based on Biochemical, Structural and Computational Analysis of two GH1 Enzymes from *Trichoderma harzianum* (submetido).

5.3 Resumos em Congresso

SOUZA, V. P.; IKEGAMI, C. M.; ARANTES, G. M.; MARANA, S. R. Thermostability and Protein Structural Networks: The Role of Hub Residues. In: 45a. Annual Meeting of Brazilian Society for Biochemistry and Molecular Biology, 2016, Natal. 45a. Annual Meeting of Brazilian Society for Biochemistry and Molecular Biology, 2016. v. 2016.

SOUZA, V. P.; IKEGAMI, C. M.; ARANTES, G. M.; MARANA, S. R. The Role of Protein Structural Network (PSN) in the thermostability. In: The 30th Annual Symposium of The Protein Society, 2016, Baltimore. The 30th Annual Symposium of The Protein Society, 2016. v. 2016.

SOUZA, V. P.; MARANA, S. R. Effects upon the stability of Sfbglu caused by mutations surrounding central residues. In: 23rd Congress of the International Union of Biochemistry and Molecular Biology (IUBMB) and 44th SBBq meeting, 2015, Foz do Iguaçu. 23rd Congress of the International Union of Biochemistry and Molecular Biology (IUBMB) and 44th SBBq meeting, 2015.

SOUZA, V. P.; MARANA, S. R. Mutations close to the central residues of the structural network of the beta glucosidase Sfbglu affect its thermostability. In: FEBS EMBO 2014, 2014, Paris. FEBS Journal Special Issue: FEBS EMBO 2014 Conference, Paris, France, 30 August-4 September 2014, 2014. v. 281. p. 630.

SOUZA, V. P.; ARANTES, G. M.; MARANA, S. R. Study of the Correlation between Structural Network and Structural and Catalytic Properties of Beta-glucosidases. In: 42ª Reunião Anual da Sociedade Brasileira de Bioquímica e Biologia Molecular, 2013, Foz do Iguaçu. 42ª Reunião Anual da Sociedade Brasileira de Bioquímica e Biologia Molecular, 2013. v. 42.

SOUZA, V. P.; FARAH, S. C.; MARANA, S. R. Crystallization of a Beta-glucosidase from *Spodoptera frugiperda*. In: 41ª Reunião Anual da Sociedade Brasileira de Bioquímica e Biologia Molecular, 2012, Foz do Iguaçu. 41ª Reunião Anual da Sociedade Brasileira de Bioquímica e Biologia Molecular, 2012. v. 41.

MARANA, S. R.; CANCADO, F. C.; **SOUZA, V. P.;** BARBOSA, J. A. R. G. Structural and molecular basis for the acidic pH optimum of digestive lysozymes from housefly *Musca domestica*. In: 36th FEBS Congress, 2011, Torino/Itália. The FEBS Journal, 2011. v. 278. p. 117-117.

SOUZA, V. P.; MARANA, S. R. Mutation of surface charged residues changes the affinity of lysozymes for bacterial walls. In: 40ª Reunião Anual da SBBq, 2011, Foz do Iguaçu. 40ª Reunião Anual da SBBq, 2011. v. 40.

SOUZA, V. P.; MARANA, S. R. Production and purification of a beta-glycosidase for crystallographic studies. In: 39ª Reunião Anual da SBBq, 2010, Foz do Iguaçu. 39ª Reunião Anual da SBBq, 2010. v. 39.