

Universidade de São Paulo
Instituto de Química

Programa de Pós-graduação em Ciências Biológicas (Bioquímica)

Felipe Akihiro Melo Otsuka

Implicações de mutações e conformações
na modulação de atividade enzimática em
uma β -glicosídeo hidrolase e na tirosina
quinase Abl1-KD

Versão corrigida da Tese defendida

São Paulo

Data de Depósito na SPG:

25/11/2022

Felipe Akihiro Melo Otsuka

Implicações de mutações e conformações
na modulação de atividade enzimática
em uma β -glicosídeo hidrolase e na
tirosina quinase Abl1-KD

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

2022

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Ficha Catalográfica

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

Bibliotecária responsável pela orientação de catalogação da publicação:
Marlene Aparecida Vieira - CRB - 8/5562

Otsuka, Felipe Akihiro Melo
O88i Implicações de mutações e conformações na modulação de
atividade enzimática em uma β -glicosídeo hidrolase e na tirosina
quinase Abl1-KD / Felipe Akihiro Melo Otsuka. - São Paulo, 2022.
101p.

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

1. β -glicosidase. 2. Homodímero. 3. Abl1. 4. Posições de
resistência. 5. Seleção de confôrmeros. I. T. II. Marana, Sandro
Roberto. orientador.

574.1925 CDD

"Implicações de mutações e conformações na modulação de atividade enzimática em uma β -glicosídeo hidrolase e na tirosinaquinase Abl1-KD"

FELIPE AKIHIRO MELO OTSUKA

Tese de Doutorado submetida ao Instituto de Química da Universidade de São Paulo como parte dos requisitos necessários à obtenção do grau de Doutor em Ciências obtido no Programa Ciências Biológicas (Bioquímica) - Área de Concentração: Bioquímica.

Prof. Dr. Sandro Roberto Marana
(Orientador e Presidente)

APROVADO(A) POR:

Profa. Dra. Deborah Schechtman
IQ - USP

Prof. Dr. Luís Eduardo Soares Netto
IB - USP

Prof. Dr. Flávio Henrique da Silva
UFSCar

SÃO PAULO
08 de fevereiro de 2023

“Change is the essential process of all existence.”

Mr. Spock; Star Trek TV series (1969)

Agradecimentos

A partir da admiração que eu tenho pela natureza, embarquei nesta jornada para compreender por que ela é tão fascinante aos meus olhos. E para minha surpresa, a fonte do meu entusiasmo vem das pessoas que tive a sorte de conhecer, conviver e crescer. Por isso, aqui demonstro a gratidão que sinto por essas pessoas que fundamentalmente construíram o meu ser e estimulam a minha admiração pelos fenômenos naturais que nos cerca.

Agradeço a minha família que me ensinaram o que é amor e respeito. Especialmente aos meus pais, Lúcio e Marise, que fazem parte das minhas doces primeiras lembranças, são meus alicerces na vida. A minha irmã Thais, a quem eu posso depositar toda a minha confiança, meu verdadeiro porto seguro. A minha avó Eliana e dinda Sônia, as quais são meu norte para uma vida honesta e de alegrias. Meus tios e tias Marcelo, Fabiana, Aroldo, Wanderly, Fausto, Leonor, Takeshi e Teresa que me transmitem imensa alegria e carinho nas suas presenças. Meus primos e primas, André, Vitória, Rodrigo, Ananda, Melissa, Milene, Marcel, Marcos, Andrea, Amanda e a recém-chegada Maria Eduarda que me encantam muito. E aqueles que não estão mais em terra, mas que guardo na minha lembrança com muito amor.

Devo um muito obrigado aos professores e mentores que me educaram moral e cientificamente. Em especial ao meu orientador de doutorado Prof. Sandro R. Marana que sou grato por instigar valores de liberdade científica e criticismo. Ao meu orientador do primeiro ano do doutorado, Prof. Roberto K. Salinas que me acolheu no programa de pós-graduação. Ao Prof. Humberto R. Matos que me ensinou a pensar de forma científica. Ao Prof. Sinisa Bjelic da Linnéuniversitetet pela mentoria científico-técnica e cultural durante o intercâmbio em Kalmar-Suécia, *tack så mycket*. E aos mestres e professores do Instituto de Química que incrementaram na minha formação de pesquisador.

Agradeço aos amigos e colegas de laboratório que tive o prazer de conhecer e de discutir sobre novas ideias e perspectiva do mundo. Em especial à Tatiana, Angie, Jose, Ignácio, Marcus, Adolfo, Philipe, Rafael, Vitor, Bárbara, Renata, Jennifer e Dick Sjöström. Também agradeço aos técnicos Camila, Ilda, Giliard, Christiane e Ivanilde pelo suporte nas atividades laboratoriais.

Agradeço a minha parceira e namorada Monalisa R. Bettim pela sua amável companhia e ricas discussões. Como também a Iara, Silvio e Tadeu que me receberam de braços abertos e fazem-me sentir como membro da família.

O meu obrigado a sociedade e instituições brasileiras que por investimentos em ensino superior e pesquisa tornou possível a minha trajetória científico-acadêmica. Agradeço as instituições de fomento à pesquisa que incluem a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), o Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e a Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) que financiou

o meu doutorado no Brasil (processo 2018/18537-4) e de estágio pesquisa no exterior (processo 2019/25955-0). Também sou grato pela infraestrutura e administração do Instituto de Química da USP e da Linnéuniversitetet.

Todas as figuras obtidas de fontes externas são usadas com permissão e citadas apropriadamente.

Resumo

Otsuka, F.A.M. **Implicações de mutações e conformações na modulação de atividade enzimática em uma β -glicosídeo hidrolase e na tirosina quinase Abl1-KD**. 2022. 101 p. Tese de Doutorado – Programa de Pós-graduação em Ciências Biológicas (Bioquímica). Instituto de Química, Universidade de São Paulo, São Paulo.

Enzimas são catalisadores de reações químicas, cuja atividade é sensível a sua sequência de aminoácidos, estado oligomérico e as condições do meio em que se encontram, *e.g.*, temperatura e força iônica. Fundamentalmente, populações estatisticamente discretas com conformações distintas influenciam e conferem estados de maior e menor atividade à enzima. Nesta tese, são discutidos em dois tipos de enzimas as interrelações entre mutações, conformações e atividade catalítica. No capítulo I, é examinada uma β -glicosídeo hidrolase de *Spodoptera frugiperda* (Sf β gly) expressa e purificada em sistema heterólogo (*E. coli*), a qual sua atividade é sensível ao seu estado oligomérico. Além disso, mutações acumulativas em sua sequência são capazes de alterar sua estereosseletividade frente a substratos sintéticos. Os resultados indicam que a Sf β gly em solução forma um equilíbrio de estado entre dímero e monômero com uma constante de dissociação (K_D) de $3,7 \pm 0,3 \mu\text{M}$, sendo a forma dimérica cerca de 2,5 vezes mais ativa comparada com seu estado monomérico quando separadas por filtração em gel. Contudo, a constante cinética de dissociação do dímero (k_{off}) estimada em $277 \cdot 10^{-6} \text{ s}^{-1}$ sugere que para alcançar o estado monomérico é preciso passar por uma elevada barreira de energia de ativação (E_a) de cerca de 18,5 kcal. A explicação deste fenômeno tem respaldo na teoria de que proteínas possuem estados conformacionais, e no caso da Sf β gly, o estado dimérico e monomérico possuiriam diferentes conformações que também estariam associadas com a atividade da enzima. Mutações na interface do dímero planejadas para desestabilizar a ligação entre as subunidades validam que a atividade do monômero é menor comparada à fração contendo dímeros da enzima selvagem. Ainda com respeito a mutações, uma variante da Sf β gly com 47 mutações acumuladas na sua sequência gerou uma enzima monomérica e com capacidade de hidrolisar substrato com ligação α -glicosídica, uma função rara entre as β -

glicosidases. No capítulo II, a temática gira em torno de mutações em uma tirosina quinase que conferem resistência a medicamentos (inibidores) para tratamento da Leucemia Mieloide Crônica (CML). O causador da CML e o alvo terapêutico é um domínio tirosina quinase da enzima Abl1-KD, que é constitutivamente ativa nesta doença. Para entender como as mutações de resistência atuam nas diferentes conformações da Abl1-KD, simulamos computacionalmente com o Rosetta o efeito estabilizador de mutações no resíduo e da estrutura no contexto da conformação. Além disso, através da inferência das sequências ancestrais de Abl1-KD, verificamos se as posições da estrutura primária mais conhecidas em conferir resistência aos inibidores sofrem algum grau de constrição evolutiva. Os resultados mostram que a energia conformacional da forma inativa estimada pelo Rosetta é $4,4 \pm 2,2$ REU (*Rosetta Energy Units*) mais estável comparada à conformação ativa, indicando que a conformação inativa é mais populada. Interessantemente, as mutações em posições que conhecidamente conferem resistência aos inibidores de tratamento da CML são capazes de remodelar a energia conformacional da Abl1-KD, tornando a conformação ativa mais estável, como no caso das mutações: M244V, G250E, Q252H, Y253H, F359VC e H396P. Em relação à energia individual dos resíduos, as mutações G250E, Q252H, Y253H, E255V, V299L, F359(C,V), e H396P estão mais estáveis na conformação ativa da Abl1-KD. Finalmente, a história evolutiva da Abl1-KD inferida por filogenia indica que as posições conhecidas por gerar resistência possuem uma menor taxa de variação de resíduo através de análise por matriz de peso score–posição (PSSM). Em conclusão, mutações na Abl1-KD capazes de remodelar sua energia conformacional estabilizando sua forma ativa, ocorrem em posições moderadamente conservadas que são responsáveis por estabilizar a forma inativa da enzima, e estão envolvidas no fenótipo de resistência observado após abrupta pressão seletiva imposta pelos inibidores de quinases.

Palavras–chave: β –glicosidase, homodímero, seleção de confôrmeros, Abl1, posições de resistência, reconstrução de ancestrais

Abstract

Otsuka, F.A.M. **Implications of mutations and conformations in the modulation of enzymatic activity in a β -glycoside hydrolase and in the tyrosine kinase Abl1-KD.** 2022. 101 p. *Ph.D. Thesis* – Graduate Program in Biochemistry. Instituto de Química, Universidade de São Paulo, São Paulo.

Enzymes that catalyze chemical reactions, whose activity is sensitive to their amino acid sequence, oligomeric state and the conditions of the medium in which they are found, *e.g.*, temperature and ionic strength. Fundamentally, statistically discrete populations with distinct conformations influence and confer states of greater and lesser activity to the enzyme. In this thesis, the interrelations between mutations, conformations and catalytic activity are discussed in two types of enzymes. In chapter I, a β -glycoside hydrolase from *Spodoptera frugiperda* (Sf β gly) expressed and purified in a heterologous system (*E. coli*) is examined, which its activity is sensitive to its oligomeric state. Furthermore, cumulative mutations in its sequence can alter its stereoselectivity against synthetic substrates. The results indicate that Sf β gly in solution forms an equilibrium of state between dimer and monomer with a dissociation constant (K_D) of 3.7 ± 0.3 μ M, the dimeric form being about 2.5 times more active compared to its monomeric state when separated by gel filtration. However, the kinetic dimer dissociation constant (k_{off}) estimated at $277 \cdot 10^{-6} \text{ s}^{-1}$ suggests that to reach the monomeric state it is necessary to pass through a high activation energy barrier (E_a) of about 18.5 kcal. The explanation of this phenomenon is supported by the theory that proteins have conformational states, and in the case of Sf β gly, the dimeric and monomeric state would have different conformations that would also be associated with the enzyme activity. Mutations in the dimer interface designed to destabilize the binding between the subunits validate that the monomer activity is lower compared to the fraction containing wild-type enzyme dimers. Still regarding mutations, a variant of Sf β gly with 47 accumulated mutations in its sequence generated a monomeric enzyme capable of hydrolyzing substrate with α -glycosidic bond, a rare function among β -glycosidases. In chapter II, the theme revolves around mutations in a tyrosine kinase that confer resistance to drugs (inhibitors) for the treatment of Chronic Myeloid Leukemia (CML). The causative agent of

CML and the therapeutic target is a tyrosine kinase domain of the enzyme Abl1-KD, which is constitutively active in this disease. To understand how resistance mutations act on the different conformations of Abl1-KD, we computationally simulated with Rosetta the stabilizing effect of mutations in the residue and structure in the context of the conformation. Furthermore, by inferring Abl1-KD ancestral sequences, we verified whether the primary structure positions best known to confer resistance to inhibitors suffer some degree of evolutionary constriction. The results show that the total conformational energies of the inactive form estimated by Rosetta is 4.4 ± 2.2 REU (Rosetta Energy Units) more stable compared to the active conformation, indicating that the inactive conformation is more populated. Interestingly, mutations in positions that are known to confer resistance to CML treatment inhibitors are capable of remodeling the conformational energy of Abl1-KD, making the active conformation more stable, as in the case of mutations: M244V, G250E, Q252H, Y253H, F359VC and H396P. Regarding the individual energy of the residues, mutations G250E, Q252H, Y253H, E255V, V299L, F359(C,V), and H396P are more stable in the active conformation of Abl1-KD. Finally, the evolutionary history of Abl1-KD inferred by phylogeny indicates that positions known to generate resistance have a lower rate of residue change via weight score–position matrix (PSSM) analysis. In conclusion, mutations in Abl1-KD capable of remodeling its conformational energy stabilizing its active form, occur in moderately conserved positions that are responsible for stabilizing the inactive form of the enzyme, and are involved in the resistance phenotype observed after abrupt selective pressure imposed by inhibitors of kinases.

Keywords: β -glycosidase, homodimer, conformer selection, Abl1, resistance positions, ancestor reconstruction

Lista de abreviaturas

Abs: absorvância

Abl1-KD: domínio quinase do gene Abl1 de humanos

Abltide: substrato peptídico para Abl1

ADP: adenosina difosfato

ATP: adenosina trifosfato

AU: unidades arbitrárias de absorvância

BCR-Abl1: oncogene da fusão entre BCR com Abl1

CAZy: banco de dados das *carbohydrate active enzymes*

cDNA: DNA complementar

CML: leucemia mieloide crônica

E_a : energia de ativação

EC: *Enzyme Commission*

EM_λ : comprimento de onda de emissão

EX_λ : comprimento de onda de excitação

F10: Tampão fosfato de potássio 10 mM pH 6,0

F100: Tampão fosfato de potássio 100 mM pH 6,0

F_{337} : fluorescência em 337 nm

f_d : fração de dímero

GH: glicosídeo hidrolase

HPLC: cromatografia líquida de alta eficiência

IEX: cromatografia de troca iônica

IPTG: isopropil β -D-1-tiogalactopiranosídeo

K_D : constante de dissociação

k_{cat} : constante catalítica

k_{cat}/K_M : constante de eficiência catalítica

K_m : constante de Michaelis-Menten

k_{obs} : constante cinética de observação

k_{off} : constante cinética de dissociação

k_{on} : constante cinética de associação

mAU: unidades arbitrárias de absorbância

MALS: espalhamento de luz estático multi-ângulo

Ni-NTA: resina de afinidade de níquel carregado

p-nf: *p*-nitrofenil

PSSM: matriz de peso score-posição

REU: unidades de energia do Rosetta

SDS-PAGE: Eletroforese desnaturante utilizando SDS

SEC: Cromatografia de gel filtração

Sfβgly: β-glicosidase de *Spodoptera frugiperda*

TK: tirosina quinase

Sumário

Introdução geral.....	17
Capítulo 1 – Atividade e oligomerização da Sfβgly	19
1.1 – <i>Introdução</i>	19
1.1.1 – Função e família de glicosídeo hidrolases	19
1.1.2 – Catálise α– e β– estereosseletiva de glicosidasas.....	21
1.1.3 – Oligomerização das glicosídeo hidrolases.....	23
1.1.4 – Conformação de proteínas e teoria de ligação por seleção de confôrmeros e ajuste induzido.....	25
1.1.5 – Propriedades enzimáticas e estruturais da Sfβly	27
1.1.6 – Objetivos e hipóteses	29
1.2 – <i>Materiais e Métodos</i>	31
1.2.1 – Reagentes, tampões e plasmídeos.....	31
1.2.2 – Expressão e purificação de Sfβgly	31
1.2.3 – Separação de dímeros e monômeros por SEC	32
1.2.4 – Atividade enzimática de Sfβgly	33
1.2.5 – SEC acoplado a MALS e detector de fluorescência	34
1.2.6 – Espectros de fluorescência intrínseca e cinética	34
1.2.7 – <i>Design</i> da variante de Sfβgly	35
1.3 – <i>Resultados e discussão</i>	35
1.3.1 – Purificação da Sfβgly recombinante e determinação do seu estado oligomérico.....	35
1.3.2 – Atividade enzimática de Sfβgly e sua relação com a oligomerização	38
1.3.3 – Cinética de homodimerização da Sfβgly e suas conformações.....	42
1.3.4 – Conexão da interface para o sítio ativo	50
1.3.5 – Mutantes de Sfβgly e suas implicações na oligomerização e atividade enzimática	54
1.3.6 – <i>Design</i> de variante de Sfβgly com capacidade de hidrolisar β– e α–substrato.....	57
1.4 – <i>Conclusões e perspectivas</i>	63
1.5 – <i>Publicação em jornal científico</i>	65
Capítulo 2 – Mutantes e evolução de Abl1-KD.....	67
2.1 – <i>Introdução</i>	67

2.1.1 – Resistência no tratamento de câncer: o caso da leucemia mieloide crônica e o oncogene BCR–Abl1	67
2.1.2 – Mutações de resistência e conformação da Abl1-KD	69
2.1.3 – Objetivos e hipóteses	71
2.2 – <i>Materiais e métodos</i>	72
2.2.1 – Modelos estruturais e conformações da Abl1-KD	72
2.2.2 – Refinamento estrutural de Abl1-KD nativo e mutado.....	72
2.2.3 – Otimização de sequência de Abl1-KD.....	73
2.2.4 – Reconstrução ancestral de Abl1-KD e <i>design</i> de ancestrais.....	73
2.2.5 – Expressão, purificação e ensaio enzimático dos ancestrais de Abl1-KD	74
2.3 – <i>Resultados e discussão</i>	76
2.3.1 – Energias das conformações e de mutações em Abl1-KD.....	76
2.3.2 – Otimização de sequência dependente de conformação da Abl1-KD	81
2.3.3 – Evolução e variabilidade de resíduos da Abl1-KD	82
2.3.4 – Atividade enzimática dos ancestrais de Abl1-KD e <i>design</i> de sequências com informações conformacionais.....	86
2.4 – <i>Conclusões e perspectivas</i>	89
2.5 – <i>Publicação em jornal científico</i>	90
Referências	91

Introdução geral

Proteínas são as biomoléculas de trabalho mais eficientes e versáteis da biologia. Um particular grupo dessas moléculas são as enzimas, verdadeiros biocatalisadores que as formas de vida empregam para que suas reações químicas ocorram com a especificidade e velocidade compatíveis à manutenção das funções celulares nesses organismos. Além disso, as enzimas são aproveitadas em processos biotecnológicos para a geração de produtos de alto valor comercial na indústria de alimentos, têxtil, farmacêutico e energia (Robinson, 2015).

Existem ao menos dois aspectos que possibilitam às enzimas serem eficientes e úteis catalisadores naturais: a) o correto enovelamento da proteína, seguido da correta disposição espacial das cadeias laterais no sítio ativo para ativar o substrato e interagir fortemente com o seu estado de transição (Knowles, 1991). b) a seleção natural que permite o refinamento da atividade enzimática durante a trajetória evolutiva de uma espécie, permitindo que a sequência de aminoácidos acumule mutações, inserções e deleções que são selecionadas e passadas para a próxima geração quando oferecem vantagens adaptativas (Singh et al., 2021).

É com base na compreensão dessas principais premissas que é possível formular conclusões sobre a atuação fisiológica de uma particular enzima em um organismo e prever como modificações na sequência de aminoácidos de uma enzima alterará a sua função. Mais empolgante ainda, é a possibilidade de utilizar esses conhecimentos para o *design* de enzimas, *i.e.*, procedimentos computacionais para criar sequências de enzimas sob medida e de função preestabelecida, seja para catalisar novas reações químicas ou para criar máquinas moleculares com potencial de revolucionar a nanotecnologia (Zhou et al., 2020).

Esta tese de doutorado utilizou-se de fundamentos físico-químicos e evolutivos para buscar responder perguntas que concernem à atividade enzimática e tem a finalidade de discorrer sobre os aspectos estruturais que modulam a função das enzimas a nível de estrutura quaternária e de estados conformacionais, além de verificar experimentalmente as implicações na atividade após estratégias

racionais e computacionais para alteração de seus aminoácidos. Para isso, duas enzimas foram estudadas e descritas em capítulos separados:

I – Enzima Sfbgly (Uniprot id: O61594), um glicosídeo hidrolase da família GH1 originalmente expressa em um inseto chamado *Spodoptera frugiperda*, também conhecida como lagarta–militar (Marana et al., 2001).

II – Enzima Abl1-KD (Uniprot id: P00519), o domínio catalítico isolado da proteína tirosina quinase humana que catalisa a transferência de grupos fosfatos para outras proteínas (Shtivelman et al., 1986).

Em cada capítulo será apresentada uma **introdução** seguida dos objetivos e hipóteses formuladas para entender como mudanças de estado oligomérico ou de sequência modulam a atividade enzimática, seguido da descrição dos **materiais e métodos** empregados para resolver e testar as hipóteses. Depois serão mostradas as evidências experimentais e as interpretações dos **resultados** obtidos. Em seguida serão apresentadas as **conclusões e perspectivas** que irão conter uma opinião sobre os fatos que resultaram dos experimentos, assim como as principais perguntas que persistem na área. Finalmente serão mostradas as **publicações em jornal científico** que foram geradas a partir desta tese.

Capítulo 1

Relação entre atividade enzimática e oligomerização da β-glicosídeo hidrolase de *Spodoptera frugiperda* (Sfβgly)

1.1 – Introdução

1.1.1 – Função e família de glicosídeo hidrolases

As enzimas conhecidas como glicosídeo hidrolase (GH) possuem papel fundamental na manutenção da vida e dos ecossistemas por atuar em substratos que são conjugados de carboidratos, considerados uma das substâncias orgânicas mais abundantes do planeta e de alto valor econômico (BeMiller, 2004; Lichtenthaler & Peters, 2004; O. Silva et al., 2018). Historicamente, as GHs foram protagonistas em diversos descobrimentos, por exemplo no desenvolvimento da teoria do modelo de “chave e fechadura” (Fischer, 1894), na síntese de glicosídeos (Bojarová-Fialová & Křen, 2007) e na indústria como propulsor de novos produtos, como por exemplo o leite sem lactose (Hari Krishna, 2002). GHs realizam a hidrólise da ligação glicosídica, gerando açúcares livres a partir de seus substratos que podem ser di-, oligo e polissacarídeos ou ainda glicoconjugados (Kötzler et al., 2014) (**Figura 1.1**). A atividade enzimática de GHs é classificada com o código EC 3.2.1.z, conforme estabelecido pela *Enzyme Commission* (EC) do *Nomenclature Committee of the International Union of Biochemistry and Molecular Biology* (<https://iubmb.org/>), onde , “3” indica que a reação é de hidrólise ($A-B + H_2O = AH + BOH$), “2” indica que hidrolisam ligações glicosídicas, “1” indica que é específica para a ligação O-glicosídica e “z” representa a especificidade do substrato. Portanto, GHs são um grupo de enzimas que disponibilizam e movem sacarídeos para diversos processos celulares críticos, como o metabolismo energético, sinalização celular, estruturação da célula e produção de precursores de ácidos nucleicos (Davies et al., 2005).

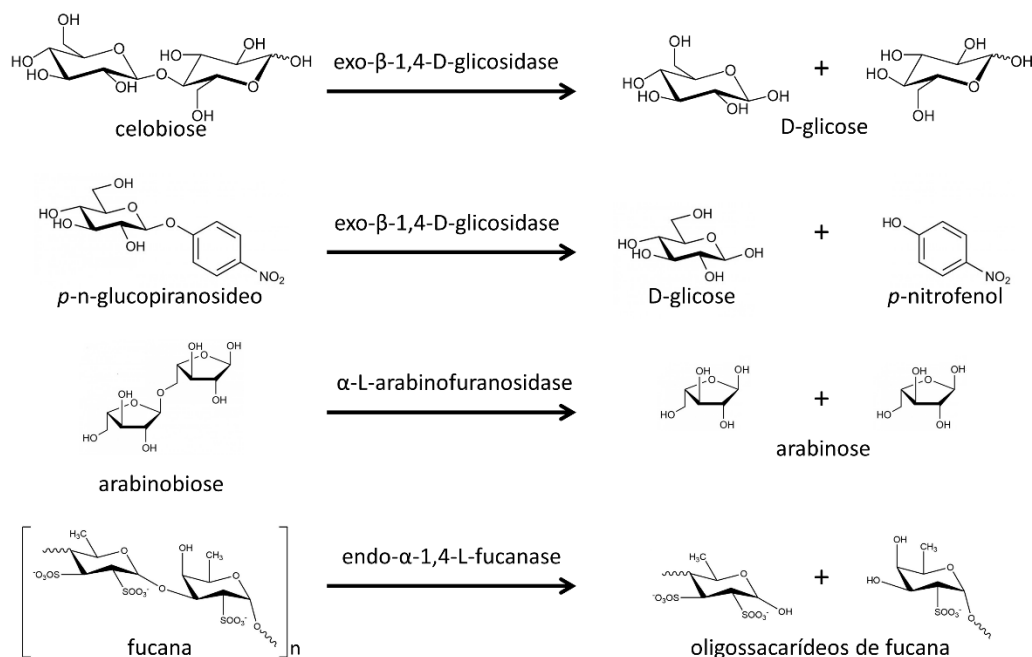


Figura 1.1 Exemplos de substratos de GHs.

A reação de catálise em GHs é estereosseletiva, do tipo ácido/base e ocorre através de dois resíduos de ácidos glutâmico ou aspártico catalíticos (**Figura 1.2**), resultando no sacarídeo livre que pode ter a posição da hidroxila no carbono anomérico que participava da ligação glicosídica sendo invertida ou retida. No caso das GHs denominadas inversoras, a própria água efetua um ataque nucleofílico e substitui um dos sacarídeos que faziam a ligação O-glicosídica através de um estado de transição tipo íon oxocarbênio, gerando, por exemplo, um α -sacarídeo como produto de um β -substrato (**Figura 1.2a**). Por outro lado, a hidrólise catalisada por GHs pode ser feita através de mecanismo de retenção, que ocorre *via* ataque nucleofílico de um dos carboxilatos catalíticos, gerando um intermediário glicosilado da enzima, enquanto o sacarídeo que é o grupo de saída é auxiliado pelo outro ácido carboxílico catalítico que cede seu próton. Depois, a água é ativada pelo segundo carboxilato, que agora age como uma base, resultando em um ataque nucleofílico no carbono anomérico do intermediário glicosilado e formação de um íon oxocarbênio como estado de transição, que irá dissociar-se resultando no sacarídeo livre (**Figura 1.2b**). Neste mesmo tipo de reação de retenção, se ao invés da água, o intermediário glicosilado for atacado por um grupo hidroxila de outro sacarídeo, que procederá como nucleófilo, a reação será chamada

de transglicosilação e haverá a formação de um novo glicosídeo (Bojarová-Fialová & Křen, 2007; Withers, 2001).

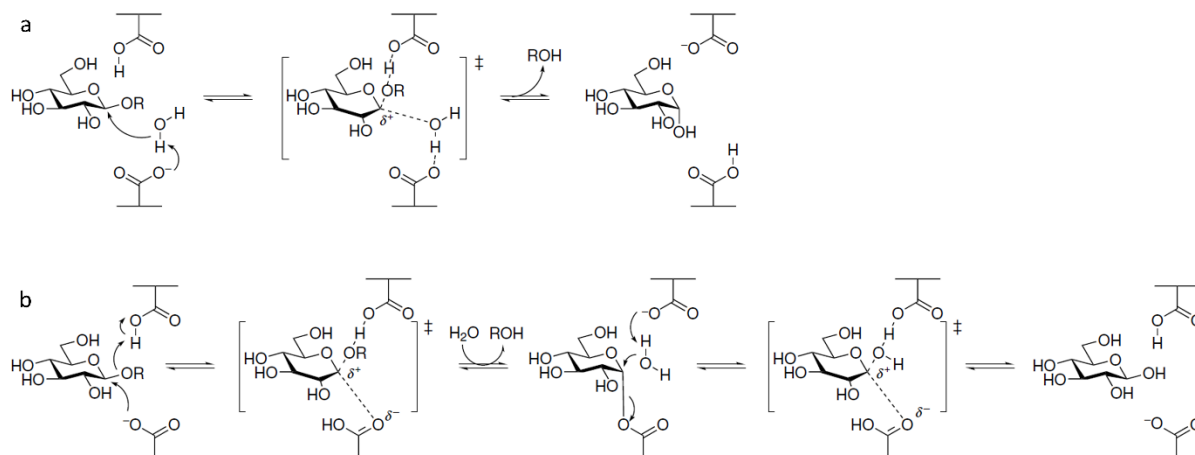


Figura 1.2. Mecanismo de reação de glicosídeo hidrolases. **a:** mecanismo de inversão (β para α) **b:** mecanismo de retenção. Retirado da referência (Bojarová-Fialová & Křen, 2007).

Diversas GHs já foram identificadas e caracterizadas em organismos de diferentes complexidades, desde bactérias até humanos (Ketudat Cairns & Esen, 2010; Nakamura et al., 2019). Para uma eficiente e organizada catalogação desse enorme grupo de enzimas, o banco de dados chamado CAZy (Carbohydrate Active Enzymes; <http://www.cazy.org/>) classifica as GHs em grupos numéricos baseados em similaridade de sequência e chamados de famílias, as quais atualmente vão da GH1 até a GH173 (Lombard et al., 2014). Por sua vez as famílias relacionadas estruturalmente ainda podem ser organizadas em grandes clãs, identificados por letras, que vão de GH-A até GH-R. As enzimas que compõem uma família compartilham do mesmo tipo de enovelamento, mecanismo de reação e tipos de resíduos catalíticos. Em termos de substratos, cada enzima poderá ser cataliticamente promíscua atuando sobre vários tipos de glicosídeos, *e.g.*, lactose, celobiose etc. No entanto, cada enzima irá ser particularmente mais ativa sobre um determinado tipo de substrato. Por exemplo, uma lactase da família GH1 (EC 3.2.1.108) terá maior atividade sobre lactose comparada à celobiose, enquanto uma β -glucosidase da família GH1 (EC 3.2.1.21) poderá ter maior atividade de celobiose comparada a sua atividade lactase (Asp et al., 1969; Marana et al., 2000).

1.1.2 – Catálise α - e β - estereosseletiva de glicosidases

A diversidade estereoquímica de glicosídeos promove funcionalidade e versatilidade na medida que monossacarídeos são ligados entre si. Um dos grandes determinantes desta complexa estereoquímica é a configuração do carbono anomérico, α ou β , envolvido na ligação entre dois monossacarídeos (**Figura 1.3**). As glicosídeo hidrolases se destacam em reconhecer o correto substrato para realizar a sua função, e em geral, se restringe à apenas um tipo de configuração α - ou β - do substrato (Thorlby et al., 2004). Existem raras GHs bifuncionais capazes de hidrolisar substratos de ambas as configurações anoméricas, como a GH4 fosfo- α -glucosidase da *Fusobacterium mortiferum* (Pikis et al., 2002), e outras β -glicosidase capazes de hidrolisar α -substratos com atividades menores a 1% relativa ao substrato de maior especificidade (Florindo et al., 2018; Seshadri et al., 2009).

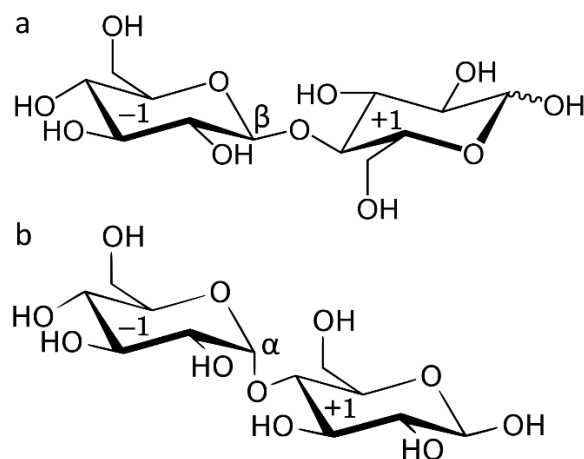


Figura 1.3. Exemplo de dois dissacarídeos que se distinguem pela configuração da ligação O-glicosídica no carbono anomérico apontado como β para a celobiose (a) e α para maltose (b). Indicações “-1” e “+1” correspondem a porção da molécula reconhecida por GHs como não redutora e redutora, respectivamente.

Ainda não há uma explicação molecular satisfatória para a discriminação entre α e β substratos de glicosídeo hidrolases, mas sabe-se que ambos os tipos de substrato compartilham equivalentes resíduos catalíticos conservados no sítio ativo, os mesmos intermediários de reação e mecanismos de hidrólise por inversão ou retenção da configuração (**Figura 1.2**). Também há registros de β -glicosidases que são inibidas por um α -substrato, ou seja, capazes de competir pelo sítio ativo (Ferreira & Terra, 1983). Possivelmente, as intrincadas redes de ligação de hidrogênio no sítio ativo com o substrato (Badieyan et al., 2012) e a correta estabilização dos

intermediários de reação (Chiba, 1997) são fatores cruciais na seleção e hidrólise do substrato de configuração apropriado.

1.1.3 – Oligomerização das glicosídeo hidrolases

O processo de oligomerização de proteínas, também conhecido como formação de estrutura quaternária, ocorre quando cadeias de proteínas isoladas se associam por interações moleculares não-covalentes, geralmente de modo específico, em muitas situações promovendo alterações de suas propriedades funcionais. Um caso clássico é a hemoglobina, cuja afinidade de ligação ao oxigênio é influenciada por sua estrutura quaternária (Pauling, 1935).

Em geral, proteínas expressas em ambiente celular formam algum tipo de oligômero, uma característica que é mais a regra do que exceção (Goodsell & Olson, 2000). É possível classificar a oligomerização de duas formas, a primeira diz respeito à identidade das proteínas que interagem. Se duas ou mais cadeias da mesma proteína, *i.e.*, subunidades, se associam, são homo-oligômeros, e se forem subunidades de cadeias diferentes, serão chamados de hetero-oligômeros. A segunda forma de classificação diz respeito à quantidade e grupo de simetria da associação entre subunidades. O grupo de simetria pode ser cíclico, dihedral, cúbico e sem simetria (Goodsell & Olson, 2000). Por exemplo, duas subunidades formando um homodímero com um eixo de simetria rotacional terá simetria cíclica C₂, e quatro subunidades formando um homotetrâmero contendo um eixo de simetria rotacional e um eixo perpendicular de simetria dupla terá simetria dihedral D₄ (**Figura 1.4**).

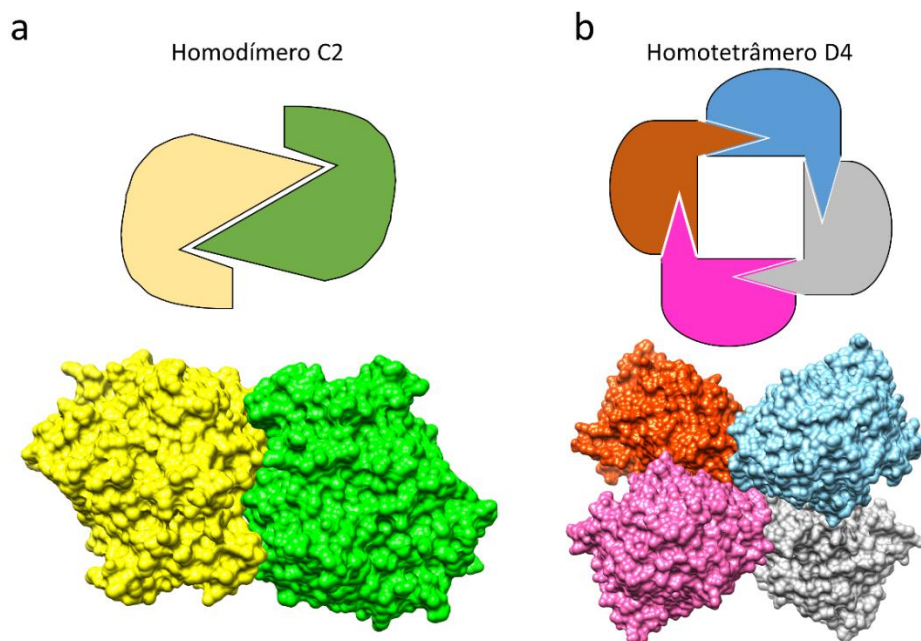


Figura 1.4. Exemplos de oligômeros. **a:** duas subunidades da mesma proteína, uma GH (PDB: 5CG0), formando um homodímero. **b:** quatro subunidades da mesma proteína, uma GH (PDB: 5DT7), formando um homotetrâmero.

Nem todas as GHs formam oligômeros, mas para aquelas que o fazem, a oligomerização parece ter alguma relação com a sua função, ou seja, possivelmente é resultado de uma adaptação evolutiva (Ali & Imperiali, 2005). É o caso de uma GH1 β -glucosidase de *Exiguobacterium antarcticum* (Figura 1.4b), que forma tetrâmeros adaptados para maior atividade em ambientes frios (Zanphorlin et al., 2016). Já em uma GH57 α -amilase de *Thermotoga maritima*, os tetrâmeros são espacialmente organizados para dificultar o acesso e a hidrólise de porções ramificadas da amilopectina, enquanto outro membro desta mesma família, também uma hidrolase de α -glucanas, ocorre estritamente como monômero, sendo capaz de atuar em substratos ramificados (Dickmanns et al., 2006; Santos et al., 2011). Finalmente, em um caso mais complexo, a β -glucosidase GH1 da aveia forma estados oligoméricos fibrilares que se organizam formando um túnel, onde localizam-se os sítios ativos desses monômeros, exibindo um incremento da atividade enzimática nesta forma oligomérica (KIM et al., 2005; Kim & Kim, 2004). Portanto, o processo de oligomerização, em especial entre subunidades idênticas, se mostra uma estratégia molecular de adaptação simples e de menor custo energético para a célula comparado à produção de novas proteínas (Danielli et al., 2020), servindo como modulador da função em GHs.

1.1.4 – Conformação de proteínas e teoria de ligação por seleção de confôrmeros e ajuste induzido

É concebido que as proteínas contêm uma sequência específica de aminoácidos que é responsável por ditar uma estrutura tridimensional bem definida, processo chamado de enovelamento, e satisfatoriamente explicado pela hipótese termodinâmica (Anfinsen, 1973). Atualmente, a melhor representação deste processo é através da superfície de energia de enovelamento com forma de funil (**Figura 1.5**), que resume componentes quantitativas energéticas e qualitativas estruturais na forma de superfície do funil (Dill et al., 2008). Desta forma, uma cadeia proteica representada no fundo do funil está na forma nativa e, portanto, funcional, enquanto no topo do funil, a proteína está em um estado de alta entropia, *i.e.*, desnaturada e inativa.

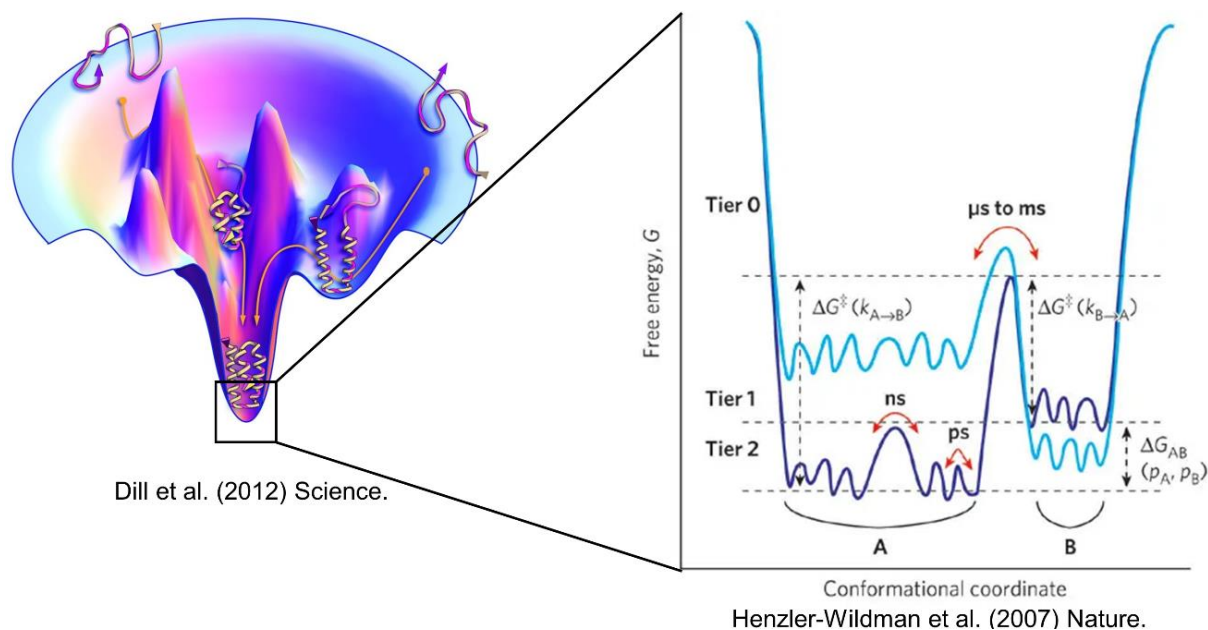


Figura 1.5. Representação da superfície de energia de enovelamento de uma proteína em forma de funil. No fundo do funil, a superfície de energia contempla as conformações (*e.g.*, “A” e “B”) que a proteína pode amostrar na sua forma nativa, incluindo as barreiras energéticas entre as conformações que ditam a sua cinética de interconversão. Figuras retiradas e modificadas de (Dill & MacCallum, 2012; Henzler-Wildman & Kern, 2007).

É importante ressaltar que a estrutura nativa da proteína é uma média de conformações que varia conforme a cinética de flutuação rotamérica nas cadeias laterais dos resíduos no interior e ao redor da proteína pela camada de solvatação de água (**Figura 1.6**) (Frauenfelder et al., 2009). Essa intrínseca dinâmica da estrutura da proteína também pode ser representada como um funil de caráter

rugoso (**Figura 1.5**) (Henzler-Wildman & Kern, 2007), onde cada conformação é separada por barreiras energéticas que ditam a cinética de troca entre conformações. Quanto mais elevada a barreira, mais lenta é a troca de uma conformação para outra. O efeito macroscópico da dinâmica da proteína e o conjunto de todas as entidades proteicas de diferentes conformações, vai por exemplo, ditar a taxa de atividade de uma enzima, considerando que distintas conformações, mesmo na forma nativa, possuem diferentes taxas de atividades (Wolf-Watz et al., 2004).

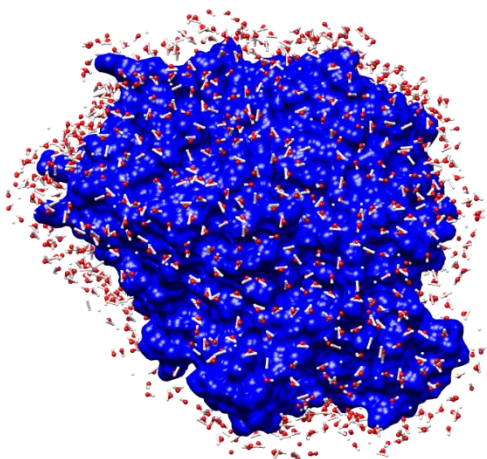
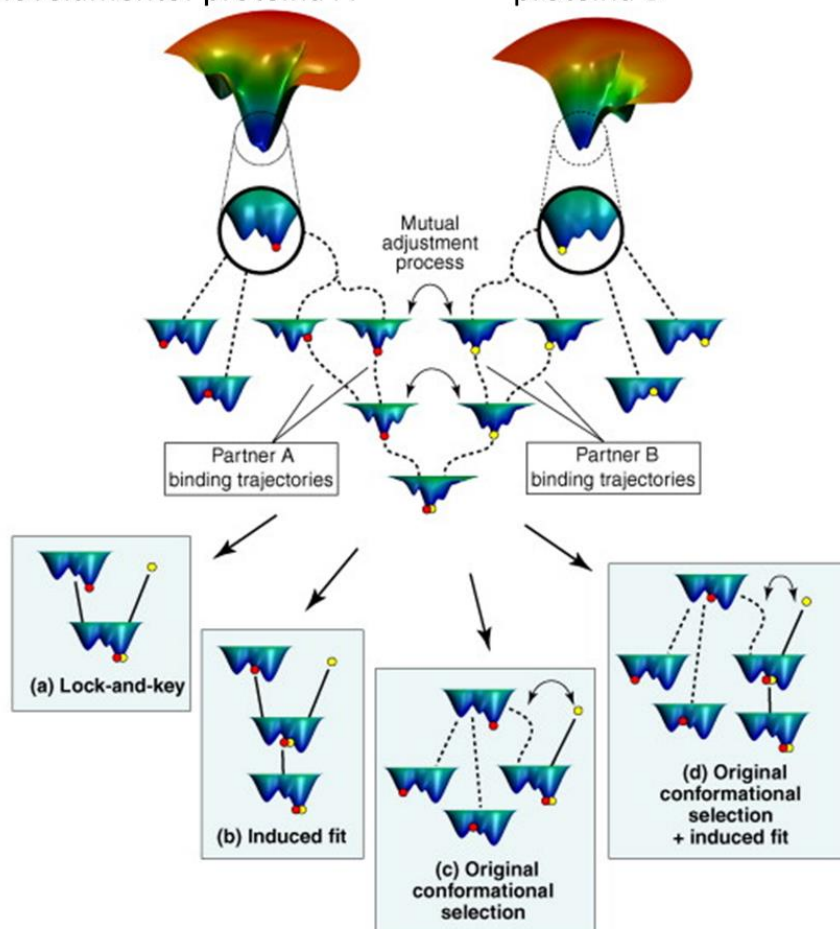


Figura 1.6. Estrutura em superfície de uma proteína (PDB 5CG0) cercada pela primeira camada de solvatação.

A alternância conformacional das proteínas dá embasamento para outra propriedade funcional muito importante: a formação de hetero e homo-oligômeros, ou seja, a interação entre proteínas. Neste contexto, considerando uma proteína “A” com populações discretas de diferentes conformações, caso uma delas tenha maior afinidade por outra molécula “B”, poderá ser observado um efeito estatístico de deslocamento populacional para esta população de maior afinidade se de fato ela encontrar e ligar com “B”, um caso denominado como seleção de confôrmeros. Em outra situação, a proteína “A” já possui certa afinidade pela outra molécula “B”, contudo se ocorre a ligação entre elas, um novo ajuste torna a afinidade de ligação ainda maior. Esses dois mecanismos de ligação, seleção de confôrmeros e ajuste induzido, entre proteínas, também são explicados por funis de superfície de energia, analogamente ao de enovelamento (Weikl & Paul, 2014). Na **Figura 1.7** é ilustrado a relação entre os diferentes mecanismos de ligação e as conformações de duas proteínas “A” e “B” representadas em posições na superfície de energia do

funil de enovelamento, como por exemplo, os cenários de seleção de conformêmeros e de ajuste induzido (Csermely et al., 2010).

Funil de enovelamento: proteína A proteína B



Csermely et al. (2010) Trends in Biochemical Sciences TIBS

Figura 1.7. Mecanismos de ligação entre proteínas em vista da superfície de energia de enovelamento. Figura retirada e modificada de Csermely et al. (2010).

1.1.5 – Propriedades enzimáticas e estruturais da Sfβgly

A Sfβgly foi originalmente caracterizada como uma celobiase associada ao glicocálix no intestino médio da lagarta *Spodoptera frugiperda*, também conhecida como lagarta-militar (Ferreira et al., 1994; Marana et al., 2000). De acordo com estes estudos anteriores, considerando que a Sfβgly era ativa sobre celobiose e oligocelodextrinas curtas (C2 a C5), foi proposto que esta enzima participaria da reação de quebra de celulose e/ou hemiceluloses presentes na dieta deste inseto. Sfβgly também possui atividade sobre glicoconjugados sintéticos (*p*-nitrofenil β-glicosídeos e metilumbeliferil β-glicosídeos) e outros naturais, como a gentiobiose e amigdalina (Marana et al., 2001).

Com base no sequenciamento do seu cDNA, sintetizado a partir do RNA extraído do intestino médio dessa lagarta, Sfbgly foi classificada pelo CAZy na família GH1 e clã GH-A, e como todos os membros desses grupos, possui uma estrutura terciária de barril $(\alpha/\beta)_8$ (**Figura 1.8**), também conhecido como TIM barrel (Banner et al., 1975). Seu cDNA também foi clonado em vetor de expressão para produção em sistema heterólogo (*E. coli* e *P. pastoris*), o que facilitou determinar as constantes cinéticas das reações por ela catalisada, sua estrutura cristalográfica (PDB 5CG0), estabilidade térmica e os efeitos de diversas mutações sítio-dirigida sobre sua especificidade pelo substrato (Marana et al., 2001, 2002; Souza et al., 2016; Tamaki et al., 2016). Esta ampla base de dados e a facilidade para sua produção heteróloga, faz de Sfbgly um ótimo modelo de estudo filogenético e enzimático de GHs. De fato, foi possível, por exemplo, identificar mutações em posições que se ligam a porção aglicona do substrato como sendo aquelas que aumentam a atividade frente à celobiose (Tamaki et al., 2016).

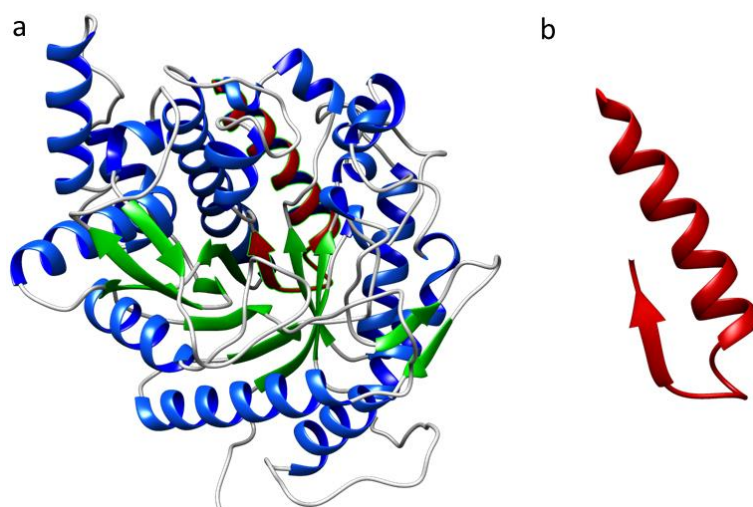


Figura 1.8. **a:** Estrutura cristalográfica da Sfbgly (PDB 5CG0) com ênfase no barril $(\alpha/\beta)_8$ e destaque das alfa-hélices e fitas beta em cores azul e verde, respectivamente. **b:** Unidade estrutural de repetição do barril $(\alpha/\beta)_8$ em vermelho.

A eficiência catalítica da Sfbgly ($k_{cat}/K_m \sim 1 \cdot 10^3 \text{ M}^{-1}\text{s}^{-1}$ para o substrato natural celobiose, Tamaki et al., 2016) não parece ser uma referência de altíssima atividade. Além de seu k_{cat}/K_m estar distante do teto teórico limitado pela difusão ($\sim 10^9 \text{ M}^{-1}\text{s}^{-1}$), é 100 vezes menor comparado a mediana da distribuição de um compilado de k_{cat}/K_m para mais de 1800 enzimas considerando apenas seus substratos naturais (Bar-Even et al., 2011). Portanto, parece que há pleno

potencial para aperfeiçoamento da atividade de Sfbgly utilizando-se de estratégias de mutações que visem estabilizar o estado ativo da enzima através de interações epistáticas distantes do sítio catalítico ou por técnicas laboratoriais de evolução dirigida. No entanto, esta tarefa não é fácil, e em geral mutações em enzimas, em especial, nas posições que fazem parte do sítio ativo, podem levar a efeitos adversos, como a formação de produtos indesejáveis, diminuição de afinidade pelo substrato e até causar a perda de atividade (Ly & Withers, 1999). Exemplo disto com a Sfbgly é que algumas mutações, mesmo distante do sítio ativo podem afetar a seletividade por substratos (Mendonça & Marana, 2011).

Sobre a estrutura quaternária da Sfbgly, é sabido que esta forma um dímero em solução (Souza, 2017). Ademais a análise deste dímero no *software* PISA (Krissinel & Henrick, 2007) revela uma interface de 905 Å² e 31 resíduos que realizam contatos reciprocamente entre as subunidades (**Figura 1.9**). A homodimerização não compromete o acesso pelo substrato até o sítio ativo, já que o túnel que dá acesso a ele está com a face voltada para a direção contrária à da interface, então não há uma modulação de função explícita pelo controle de captura do substrato. No entanto, ainda é possível que efeitos alostéricos e de estados conformacionais atuem na atividade enzimática em resposta ao estado oligomérico da enzima, fenômenos que são pouco explorados na família de GH, mas presentes e caracterizados em outras enzimas, como na superfamília de enzimas desidrogenase/reductase (Li et al., 2022).

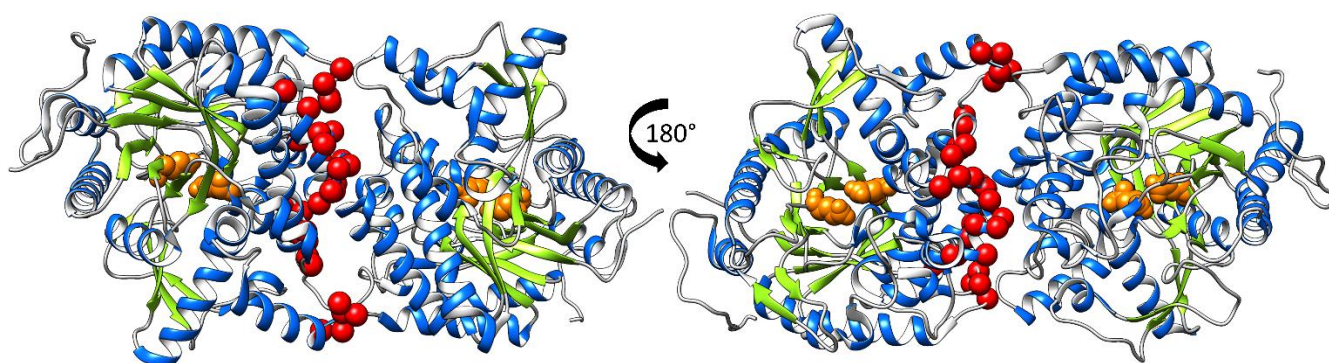


Figura 1.9. Estrutura cristalográfica do homodímero da Sfbgly (PDB 5CG0). Em esferas laranjas estão os resíduos catalíticos E187 e E399, e as esferas vermelhas, apenas na cadeia da esquerda, são os carbonos alfa de resíduos que fazem contato na interface.

1.1.6 – Objetivos e hipóteses

Frente à falta de evidências experimentais sobre a funcionalidade do estado dimérico da Sfbgly, a presente tese expõe um conjunto de experimentos e linhas de pensamento buscando mostrar que a dimerização e atividade enzimática da Sfbgly estão associadas, sugerindo que a relação desses atributos não são produtos de mera chance. Estratégias de mutação sítio dirigida serão especialmente úteis para testar hipóteses de modificar a dimerização e atividade enzimática. As principais hipóteses foram testadas da seguinte forma:

- I- A partir da Sfbgly recombinante purificada, seu estado oligomérico predominante poderá ser dímero ou monômero a depender de fatores físicos extensivos, como a concentração de proteína e força iônica da solução. Então, quanto mais concentrada a enzima, maior a chance de duas subunidades se encontrarem, enquanto a força iônica do tampão dita a intensidade entre as forças eletrostáticas dos resíduos que estão na interface do dímero. Portanto, a enzima foi purificada e analisada usando cromatografia de filtração em gel analítica (SEC) acoplado a detector de MALS e fluorescência para investigar a influência da concentração de enzima no estado oligomérico.
- II- Os estados oligoméricos e atividade enzimática de Sfbgly se relacionam através das suas conformações, *i.e.*, distintos perfis estruturais com diferentes energias e características. É pressuposto que alguma conformação da Sfbgly tenha maior afinidade à formação do dímero e seja mais ativa, enquanto outra conformação seja energeticamente estável como monômero e menos ativa. Para testar esta hipótese, foram obtidos os parâmetros de cinética enzimática (K_M e k_{cat}) das frações de dímero e monômero separadas por SEC. Além da cinética de conversão do dímero para monômero com experimentos de diluição monitorados por fluorescência intrínseca que servem para esclarecer a cinética de ligação entre as subunidades com modelos matemáticos que se aproximam da realidade física.
- III- Mutações racionalmente planejadas ou calculadas por simulações computacionais são capazes de guiar propriedades de uma enzima para uma determinada finalidade, *e.g.*, aumentar a eficiência catalítica e estabilizar uma forma oligomérica. Portanto, mutações pontuais na interface do dímero foram inseridas para estabilizar e desestabilizar o dímero e validar se de fato

o monômero é menos ativo comparado à enzima selvagem. Em outra medida mais agressiva, uma variante com 45 mutações acumuladas na sequência de Sfβgly foi desenhada utilizando *softwares* especializados, para verificar os impactos na sua dimerização e atividade.

1.2 – Materiais e Métodos

1.2.1 – Reagentes, tampões e plasmídeos

O vetor pET-46 EK-LIC (Merck Millipore, Billerica, MA, EUA) contendo a o segmento de DNA que codifica a Sfβgly (pET46-Sfβgly) já estava disponível em nosso grupo de pesquisa (Souza et al., 2016). Padrões de peso molecular para SDS-PAGE de ampla gama foram adquiridos da Bio-Rad (Hercules, CA, EUA). Tampão F100 significa tampão fosfato de potássio 100 mM, pH 6, o tampão de lise é fosfato de sódio 50 mM pH 7 contendo NaCl 300 mM e imidazol 20 mM. O tampão de eluição é fosfato de sódio 50 mM pH 7 contendo NaCl 300 mM e imidazol 300 mM. O tampão de ligação da troca iônica (IEX) é HEPES 50 mM pH 7 NaCl 50 mM. O tampão de eluição IEX é HEPES 50 mM pH 7 NaCl 1 M.

1.2.2 – Expressão e purificação de Sfβgly

O segmento de DNA que codifica a Sfβgly selvagem inserida em plasmídeo de expressão pET-46 com His₆-tag na sua porção N-terminal (pET46_His₆-Sfβgly) foi transformado em *E. coli* BL21 (DE3) competente e uma única colônia foi inoculada em 5 ml de LB com glicose 1% e carbenicilina 50 µg/ml e incubada sob agitação a 200 rpm, 35 °C por 16 h. Em seguida, 1 ml deste pré-inóculo foi adicionado a 500 ml de LB com carbenicilina 50 µg/ml. O inóculo foi incubado a 35 °C sob agitação (200 rpm) até atingir DO₆₀₀ de 0,5 – 0,7. Depois disso, isopropil β-D-1-tiogalactopiranosídeo (IPTG) foi adicionado a uma concentração final de 0,8 mM para induzir a expressão de Sfβgly por 16 h a 22,5 °C e agitação a 200 rpm. Em seguida, as células foram centrifugadas a 4.000 rpm por 20 min a 4 °C e armazenadas a -20 °C para processamento adicional.

As células peletizadas foram suspensas em tampão de lise (2:1 p/v) contendo lisozima (1 mg/ml). A lise celular foi feita por sonicação (Branson Sonifier 250, Soni-tech) usando 5 pulsos a 10% de amplitude por 10 s com intervalo de 1 min

entre os pulsos. As amostras foram mantidas no gelo durante a sonicação. Os detritos celulares foram sedimentados por centrifugação a 9.000 *g* (30 min, 4 °C). O sobrenadante foi incubado com resina Ni-NTA (Qiagen) em uma proporção de 1 ml de sobrenadante para 200 µl de resina que foi pré-equilibrada com tampão de ligação de lise. Esta suspensão foi incubada por 1 h com agitação de ponta a ponta a 8 °C. As proteínas não ligadas foram removidas por centrifugação (5.000 *g*, 1 min, 4 °C). Em seguida, as proteínas fracamente ligadas foram removidas da resina por 5 ciclos de lavagem com tampão de lise e centrifugação. Em seguida, a mistura foi eluída em 3 ciclos de incubação com tampão de eluição (120 µl de tampão de eluição para cada 200 µl de resina) por 10 min a 8 °C sob agitação suave seguido de centrifugação e coleta do sobrenadante em cada ciclo. O sobrenadante contendo o Sfβgly foi submetido à troca de tampão usando uma coluna HiTrap Dessalting (GE Healthcare) pré-equilibrada com tampão de ligação IEX. A cromatografia IEX foi realizada num sistema ÄKTAFPLC (GE Healthcare) com uma coluna MonoQ HR5/50 GL (GE) a 25 °C e um fluxo de 1 ml/min. Após a troca do tampão (ver acima), a amostra completa foi carregada na coluna. As proteínas não ligadas foram removidas por lavagem com tampão de ligação IEX (5 ml). As proteínas ligadas foram eluídas em um gradiente linear de 0 a 70% de tampão de eluição IEX em 20 volumes de coluna. As frações foram coletadas em placas de 96 poços profundos e aquelas contendo Sfβgly foram reunidas em um dispositivo de filtro centrífugo (poro de 10 kDa; Amicon Millipore). Após ciclos repetitivos de lavagem com tampão F100, a Sfβgly foi concentrada para 93,5 µM. Todas as diluições adicionais foram feitas a partir desta solução estoque. A avaliação da purificação da proteína recombinante foi feita por SDS-PAGE (LAEMMLI, 1970). O método BCA (Smith et al., 1985) foi usado para quantificar a concentração de proteína purificada e a albumina de soro bovino (Sigma-Aldrich Inc., St. Louis, MO) foi utilizada como padrão. As designações [Sfβgly] e “proteína total” são sinônimos.

1.2.3 – Separação de dímeros e monômeros por SEC

A cromatografia de filtração em gel analítica (SEC, do inglês *size exclusion chromatography*) foi realizada utilizando um sistema ÄKTAFPLC (GE

HealthCare). Duas colunas Superdex 200 HR10/300 GL (GE HealthCare) acopladas em linha ou, quando mencionado de outra forma, uma única Superdex 200 Increase 10/300 GL (GE HealthCare), foram equilibradas com tampão F100. As colunas foram imersas em gelo moído para melhorar a resolução entre os picos (Hong et al., 2012), conseqüentemente, a pressão do sistema aumentou até estabilização em 1 MPa a 0,25 ml/min devido à alta viscosidade do tampão associada à baixa temperatura. Diferentes concentrações de Sfbgly preparadas em F100 foram incubadas por pelo menos 16h na temperatura de 5 °C para alcançar o equilíbrio entre dímero e monômero antes da separação por SEC. Cada amostra foi carregada nas colunas usando um *loop* de amostra de 200 µL. As frações (0,5 ml) de dímero e monômeros separadas pelo SEC foram coletadas em uma placa de 96 poços profundos previamente congelada e contendo água congelada nos poços vizinhos aos que foram utilizados na coleta das frações contendo enzima. Os dados foram coletados e processados por meio do *software* UNICORN versão 5.11 (GE HealthCare).

1.2.4 – Atividade enzimática de Sfbgly

A velocidade inicial (v_0) de hidrólise de pelo menos 10 concentrações diferentes dos substratos: *p*-nitrofenil-β-D-glucosídeo (*p*-nf β-glucosídeo), *p*-nitrofenil-β-D-fucosídeo (*p*-nf β-fucosídeo), *p*-nitrofenil-β-D-galactosídeo (*p*-nf β-galactosídeo) e celobiose, preparados em tampão F100. Para determinar os parâmetros cinéticos (k_{cat} e K_M) foi assumindo uma cinética de estado estacionário e utilizado o modelo de cinética enzimática de Michaelis-Mentem. No caso de ensaios de atividade após SEC, apenas o substrato *p*-nf β-glucosídeo foi utilizado, e as frações diméricas e monoméricas de Sfbgly coletadas na placa gelada foram rapidamente diluídas separadamente até a concentração de 0,1 µM usando tampão F100 gelado. Cerca de 3 minutos se passaram entre o ensaio e a conclusão da SEC. A enzima diluída (50 µl) foi misturada com o substrato (50 µl) em uma placa de 96 poços em gelo e em seguida transferida para um banho-maria a 30 °C. As reações foram paradas adicionando 100 µl de carbonato de sódio 0,5 M em 4 tempos diferentes para calcular a v_0 de formação de produto ao longo do tempo, ou seja, a velocidade de reação. Depois, o produto da reação de hidrólise dos substratos

sintéticos, *p*-nitrofenol, foi quantificado através da absorbância a 415 nm, determinadas com um leitor de placas. O produto de hidrólise da celobiose é a *D*-glicose que foi quantificada utilizando kit da Vida Biotecnologia através de método enzimático.

1.2.5 – SEC acoplado a MALS e detector de fluorescência

A determinação da massa molar média das proteínas foi realizada utilizando o mesmo sistema SEC descrito acima conectado a um espalhamento de luz estático multi-ângulo miniDAWN TREOS e um detector Optilab T-rEX (Wyatt Technology Corporation, Santa Barbara CA, EUA). As condições experimentais também foram as mesmas já descritas acima. Os dados foram coletados e processados usando o *software* ASTRA versão 7.1.1.3 (Wyatt Technologies), por exemplo, para calcular a massa molar média. A SEC também foi realizada em um sistema de HPLC (Shimadzu 20AD com um detector de arranjo de diodos (PDA) e um detector de fluorescência RF-20A) usando as mesmas condições descritas acima. A eluição de Sfbgly foi avaliada em dois canais de absorbância (280 nm e 295 nm) e no modo de fluorescência (comprimento de onda de excitação (EX_{λ}) = 295 nm, comprimento de onda de emissão (EM_{λ}) = 337 nm).

1.2.6 – Espectros de fluorescência intrínseca e cinética

A fluorescência intrínseca consiste em incidir um feixe de luz com comprimento de onda conhecido na amostra contendo a enzima. Desta forma, a luz que interage com resíduos aromáticos (em especial o triptofano) os excita para que “devolvam” fótons com comprimento de onda maior. Foi utilizado um espectrofluorímetro F-7000 (Hitachi, Tóquio, Japão) para detectar a fluorescência intrínseca de amostras com Sfbgly. As larguras das fendas foram ajustadas para 1 nm e 2,5 nm para excitação e emissão, respectivamente. A temperatura entre 5 e 20 °C foi controlada conectando o suporte da cubeta a um banho térmico. Para espectros de fluorescência, comprimento de onda de excitação (EX_{λ}) foi ajustado para 295 nm e comprimento de onda de emissão (EM_{λ}) foi registrado entre 310 e 400 nm. Para varredura de fluorescência no tempo, EX_{λ} e EM_{λ} foram ajustados para 295 e 337 nm, respectivamente, e registrados ao longo do tempo. Amostras

de Sfbgly (0,5 a 93,5 μM) foram previamente incubadas por 16 h na respectiva temperatura de ensaio. As amostras também foram centrifugadas a 9.000 g por 2 min. Para a cinética da dissociação do homodímero, a amostra de Sfbgly (93 μM ou 75 μM) foi rapidamente diluída para 0,5 μM com o mesmo tampão de amostra termicamente equilibrada e, em seguida, rapidamente transferida para o espectrofotômetro de fluorescência para acompanhar a intensidade de EM_{λ} a 337 nm ao longo do tempo (18.000 s de duração com leituras cada 5 s).

1.2.7 – *Design* da variante de Sfbgly

A estrutura cristalográfica de Sfbgly (PDB 5CG0) foi baixada, a cadeia “A” foi isolada das demais e foram removidos ligantes e moléculas de água da estrutura. Depois o Rosetta (Leaver-Fay et al., 2011) foi utilizado com RosettaScripts (Fleishman et al., 2011) para adicionar hidrogênios nos aminoácidos da proteína, seguido de relaxamento com o protocolo FastRelax (Khatib et al., 2011). A estrutura relaxada foi submetida no *webserver* PROSS (<https://pross.weizmann.ac.il/step/pross-terms/>) (Goldenzweig et al., 2016), parâmetros de *design* foram proibidos nos resíduos essenciais para a catálise descritos por Tamaki (Tamaki et al., 2016) e nos que compõem a interface. Uma variante com 43 mutações foi escolhida e submetida para o próximo *round* de *design* no sítio ativo através do *webserver* FuncLib (<https://funclib.weizmann.ac.il/bin/steps>) (Khersonsky et al., 2018). Uma das sequências com 4 mutações, *i.e.*, 47 mutações em relação à selvagem, foi escolhida para ser sintetizada (GenScript, USA) e inserida em *frame* no plasmídeo de expressão pET-28a(+) e purificada como mencionado em sessões anteriores.

1.3 – Resultados e discussão

1.3.1 – Purificação da Sfbgly recombinante e determinação do seu estado oligomérico

A Sfbgly foi purificada através de estratégias usuais de purificação de proteínas. Primeiramente, foi utilizada cromatografia por afinidade com uma resina de Ni-NTA para isolar a enzima das demais no lisado bacteriano (raia 5 e 6 da **Figura 1.10a**). No entanto, dada a presença de proteínas contaminantes

minoritárias, uma nova etapa de purificação foi necessária, sendo que então foi usada cromatografia de troca iônica em uma coluna Mono Q™ (Cytiva, EUA). Deste modo, foi obtida (raia 7 da **Figura 1.10a,b**) uma amostra de pureza desejável para os testes biofísicos e enzimáticos.

Para avaliar o estado oligomérico da Sfβgly, a enzima foi concentrada até 30 μM e analisada por SEC-MALS, onde determinou-se que Sfβgly ocorre como homodímero (M_w teórico = 113,6 g/mol) com $M_w = 117 \pm 1$ g/mol e monômero (M_w teórico = 56,8 g/mol) com $M_w = 58,6 \pm 0,7$ g/mol (**Figura 1.10c**). Além disso, também foi demonstrada a sensibilidade da oligomerização à força iônica do tampão (**Figura 1.10d**), sendo que em baixa força iônica (tampão 10 mM fosfato pH 6) o estado monomérico é preponderante, enquanto em maior força iônica (tampão 100 mM fosfato pH 6) a enzima tem preferência pelo estado dimérico. É conhecido que a oligomerização de proteínas pode ser um processo reversível e que depende da força iônica do tampão (Batoulis et al., 2016; Bhattacharya et al., 2014; Kumari & Yadav, 2019; Mahatabuddin et al., 2017).

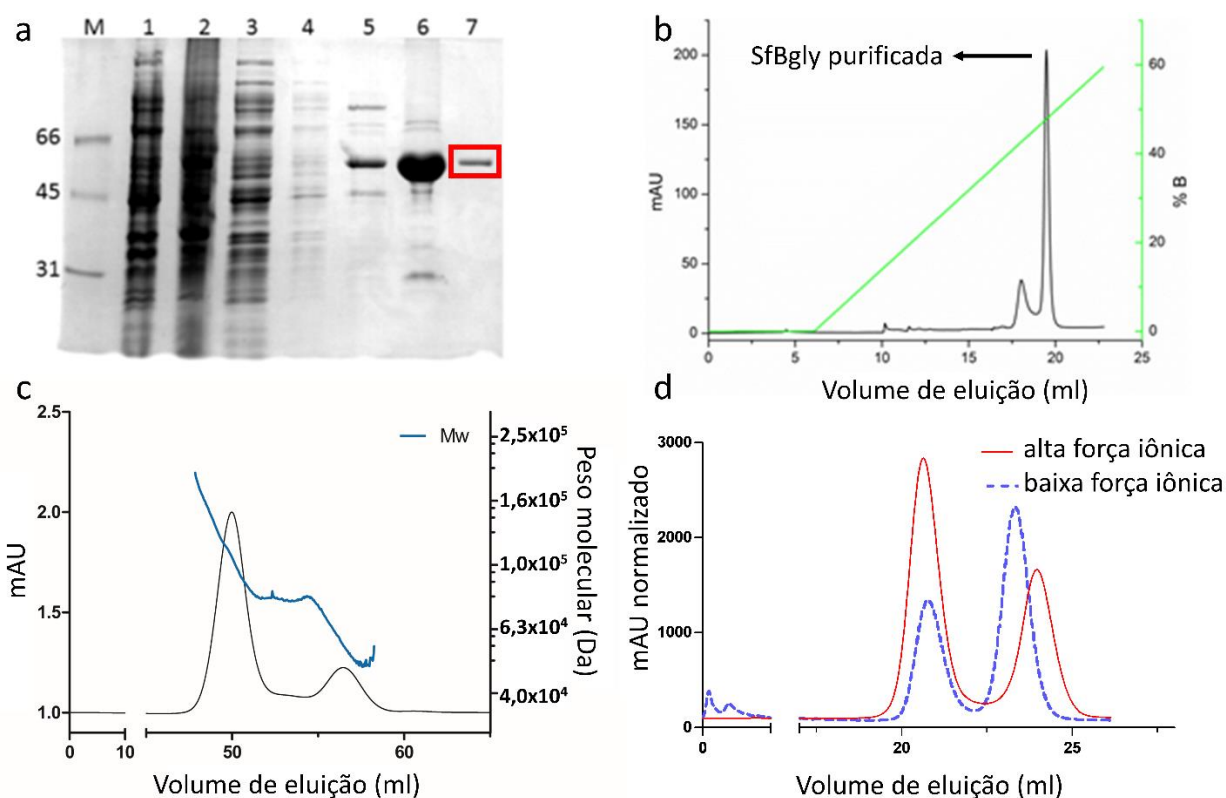


Figura 1.10. Purificação da Sfβgly recombinante expressa em bactérias BL21(DE3). **a**: SDS-PAGE das etapas de purificação; M – marcador de peso molecular (Da), 1 – cultura bacteriana não induzida, 2 – sobrenadante do lisado das bactérias induzidas, 3 – proteínas não ligadas à resina Ni-NTA, 4 – lavado usando o tampão de amostra, 5 – proteínas ligadas especificamente na resina e eluídas usando o tampão de eluição, 6 – proteínas da

fração 5 concentrada, 7 – fração coletada após a cromatografia de troca iônica com uma coluna Mono Q™ (Cytiva, EUA). **b**: Cromatografia de troca iônica da amostra (fração 5) eluída da resina Ni-NTA, %B = porcentagem de tampão de eluição contendo 1 M NaCl. **c**: 200 μl de 30 μM Sfbgly purificada injetada em sistema SEC-MALS com coluna SuperDex™ 75 10/300 (GE Healthcare, SE) a um fluxo de 0,3 ml/min. **d**: SEC da Sfbgly em tampão fosfato de potássio pH 6 de baixa (10 mM fosfato de potássio) e alta (100 mM fosfato de potássio) força iônica.

Sendo que a Sfbgly em alta força iônica apresenta-se majoritariamente como dímero, escolhemos esta condição para investigar a influência das diferentes concentrações de proteína (93,5 – 1 μM) sobre a fração de homodímero (f_d) na solução, usando novamente SEC como ferramenta analítica. Na **Figura 1.11** é mostrado que diferentes concentrações da enzima alcançam um equilíbrio entre o estado dimérico e monomérico seguindo um perfil hiperbólico. Considerando o equilíbrio $M + M \rightleftharpoons D$, a constante de dissociação resultante para este sistema a 5 °C é de $K_D = 3,7 \pm 0,3 \mu\text{M}$ ($R^2 = 0.994$). Nota-se também que mesmo em situações de altas concentrações (cerca de 25 vezes o K_D), uma pequena fração, pelo menos 10%, de monômero persiste na solução. Além disso, o caráter hiperbólico do gráfico de f_d em função de [Sfbgly] remete a um processo não-cooperativo.

Analisando o dímero no *software* PDBePISA (Krissinel & Henrick, 2007) nota-se que formação do dímero possui uma variação de energia livre de solvatação de -17 kcal/mol e, portanto, é termodinamicamente favorável. No entanto, a contínua presença do monômero em solução sugere que a cinética da oligomerização controla a dimerização e limita que todo monômero livre seja convertido em dímero.

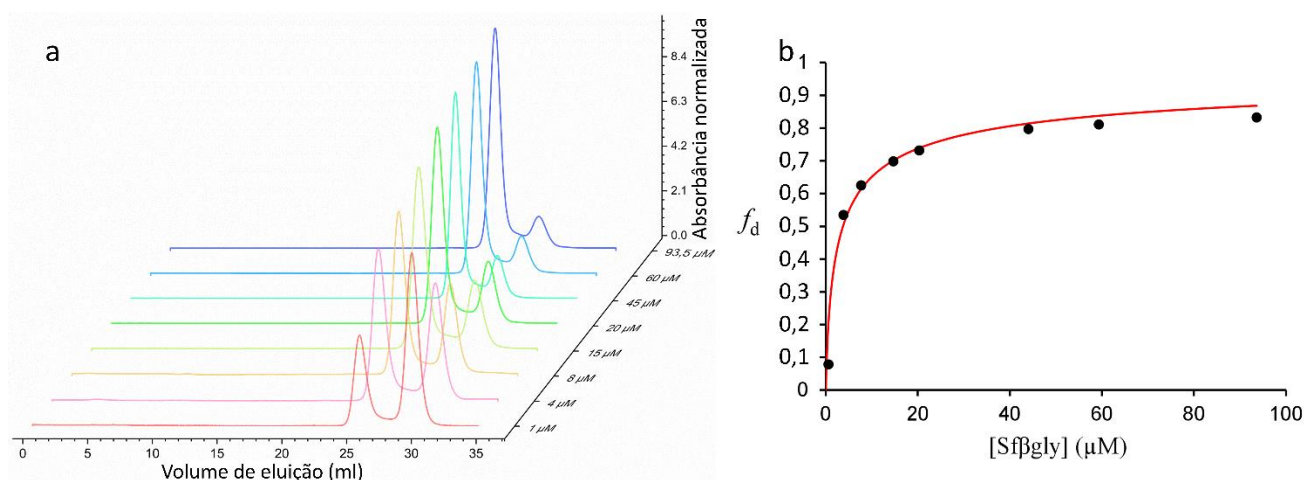


Figura 1.11. **a**: SEC da Sfbgly em oito diferentes concentrações (1 – 93,5 μM) a 5 °C. **b**: Efeito da concentração de enzima na fração de homodímero (f_d).

1.3.2 – Atividade enzimática de Sfβgly e sua relação com a oligomerização

A atividade de glicosídeo hidrolase pode ser verificada pela detecção de produtos da hidrólise de seus substratos naturais ou de substratos glicoconjugados sintéticos (**Figura 1.1**) que possuem uma *probe* que absorve luz ou emite fluorescência para facilitar sua detecção (Terra et al., 1979). Para verificar a atividade hidrolase da Sfβgly recombinante, primeiramente foram usados quatro substratos diferentes a fim de obter seus os parâmetros de cinética enzimática de estado estacionário através dos modelos e considerações de Briggs–Haldane e Henri–Michaelis–Menten (Irwin H. Segel, 1993). Na **Figura 1.12** e **Tabela 1.1** são mostrados os parâmetros cinéticos e as eficiências catalíticas para cada substrato testado, sendo um natural (celobiose) e três sintéticos (*p*-nitrofenil β-glucosídeo, *p*-nitrofenil β-fucosídeo e *p*-nitrofenil β-galactosídeo). Com relação à atividade obtida sobre celobiose, verifica-se que a eficiência catalítica (k_{cat}/K_M) da Sfβgly é de $0,71 \pm 0,08 \text{ mM}^{-1}\text{s}^{-1}$, valor similar em ordem de grandeza a estudos anteriores, *e.g.*, $k_{cat}/K_M = 1,11 \text{ mM}^{-1}\text{s}^{-1}$ (Tamaki et al., 2016) e $k_{cat}/K_M = 0,18 \pm 0,05 \text{ mM}^{-1}\text{s}^{-1}$ (Marana et al., 2002). O mesmo grau de semelhança ocorre com os substratos sintéticos. Para *p*-nitrofenil β-glucosídeo foi obtido um $k_{cat}/K_M = 0,8 \pm 0,1 \text{ mM}^{-1}\text{s}^{-1}$, enquanto que estudos anteriores determinaram $k_{cat}/K_M = 1,09 \pm 0,04 \text{ mM}^{-1}\text{s}^{-1}$ (Marana et al., 2002) e $k_{cat}/K_M = 0,26 \pm 0,01 \text{ mM}^{-1}\text{s}^{-1}$ (Mendonça & Marana, 2011). Já para o *p*-nitrofenil β-fucosídeo foi aqui obtido um $k_{cat}/K_M = 2,6 \pm 0,3 \text{ mM}^{-1}\text{s}^{-1}$, sendo que anteriormente demonstrou-se $k_{cat}/K_M = 2,8 \pm 0,9 \text{ mM}^{-1}\text{s}^{-1}$ (Marana et al., 2001) e $k_{cat}/K_M = 1,01 \pm 0,04 \text{ mM}^{-1}\text{s}^{-1}$ (Mendonça & Marana, 2011). Finalmente para o *p*-nitrofenil β-galactosídeo foi obtido um $k_{cat}/K_M = 0,04 \pm 0,01 \text{ mM}^{-1}\text{s}^{-1}$, sendo que trabalhos anteriores determinaram $k_{cat}/K_M = 0,09 \pm 0,01 \text{ mM}^{-1}\text{s}^{-1}$ (Mendonça & Marana, 2008) e $k_{cat}/K_M = 0,17 \pm 0,02 \text{ mM}^{-1}\text{s}^{-1}$ (Marana et al., 2002).

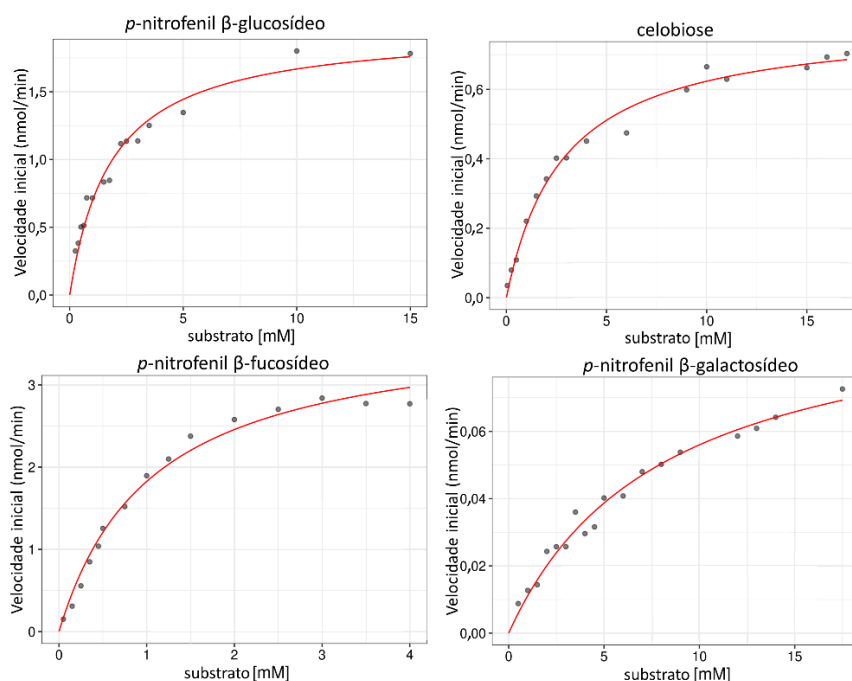


Figura 1.12. Determinação de parâmetros de cinética enzimática da Sfβgly purificada com diferentes substratos. Linha contínua é o ajuste da equação de Michaelis–Mentem sobre os dados.

Tabela 1.1. Parâmetros de cinética enzimática de Sfβgly purificada com substratos sintéticos e natural.

	K_M (mM)	V_{max} (nmol·min ⁻¹)	k_{cat} (s ⁻¹)	k_{cat}/K_M
<i>p</i> -nitrofenil β-glucosídeo	1,8 ± 0,2	1,9 ± 0,1	1,4 ± 0,1	0,8 ± 0,1
<i>p</i> -nitrofenil β-fucosídeo	1,0 ± 0,1	3,7 ± 0,1	2,8 ± 0,1	2,6 ± 0,3
<i>p</i> -nitrofenil β-galactosídeo	8,1 ± 0,9	0,10 ± 0,01	0,33 ± 0,03	0,04 ± 0,01
celobiose	2,8 ± 0,2	0,80 ± 0,02	2,02 ± 0,05	0,7 ± 0,1

Dados são média ± SEM.

Considerando o conjunto de substratos sintéticos e tomando o k_{cat}/K_M por base, temos que a especificidade relativa de Sfβgly pelo substrato seria 100% para *p*-nitrofenil β-fucosídeo, 30% para *p*-nitrofenil β-glucosídeo e 1.5% para *p*-nitrofenil β-galactosídeo, ou seja, a enzima exibe uma preferência por fucose na posição de glicone do substrato, enquanto galactose é o glicone de menor especificidade relativa. Tal perfil de especificidade pelo substrato já havia sido observado com a Sfβgly nativa e recombinante (Marana et al., 2001, 2002),

confirmando que a Sfβgly recombinante aqui produzida exhibe as propriedades enzimáticas esperadas.

Para avaliar se a atividade da Sfβgly é modulada pelo estado oligomérico, decidiu-se por usar o *p*-nitrofenil β-glucosídeo como substrato em uma série de ensaios antes e após a enzima ter seus estados oligoméricos separados por SEC. Foi predito que se um particular estado oligomérico fosse mais ativo que o outro, então a solução equilibrada contendo os dois estados ($M + M \rightleftharpoons D$) deveria ser uma média ponderada das atividades de cada estado presente na solução. Na **Tabela 1.2** e **Figura 1.13** são mostrados os parâmetros cinéticos de Sfβgly antes da separação por SEC e depois nas suas respectivas frações. Os resultados mostram que a fração contendo o homodímero de Sfβgly possui uma média de $k_{cat}/K_M = 1,03 \pm 0,04 \text{ mM}^{-1}\text{s}^{-1}$, enquanto a fração monomérica possui $k_{cat}/K_M = 0,36 \pm 0,03 \text{ mM}^{-1}\text{s}^{-1}$. Portanto, o dímero de Sfβgly é cerca de 2,5 vezes mais ativo que o monômero.

Tabela 1.2. Parâmetros cinéticos da Sfβgly purificada com *p*-nitrofenil β-glucosídeo antes da separação por SEC e das frações posteriores à separação contendo diferentes estados oligoméricos.

	Sfβgly antes da SEC	Sfβgly após SEC	
		dímero	monômero
K_M (mM ⁻¹)	1.35 ± 0.07 *	1.03 ± 0.03 *	1.9 ± 0.1
k_{cat} (s ⁻¹)	1.23 ± 0.01 *	1.06 ± 0.01 *	0.59 ± 0.01
k_{cat}/K_M (mM ⁻¹ s ⁻¹)	0.91 ± 0.06	1.03 ± 0.04 **	0.36 ± 0.03

Dados são média ± SD (n = 4 medidas independentes). Diferença estatística ao respectivo parâmetro do grupo monômero quando * para $p < .001$, e ** para $p < .01$. Análise estatística foi feita usando one way-ANOVA com pós-teste de Bonferroni entre todos os pares.

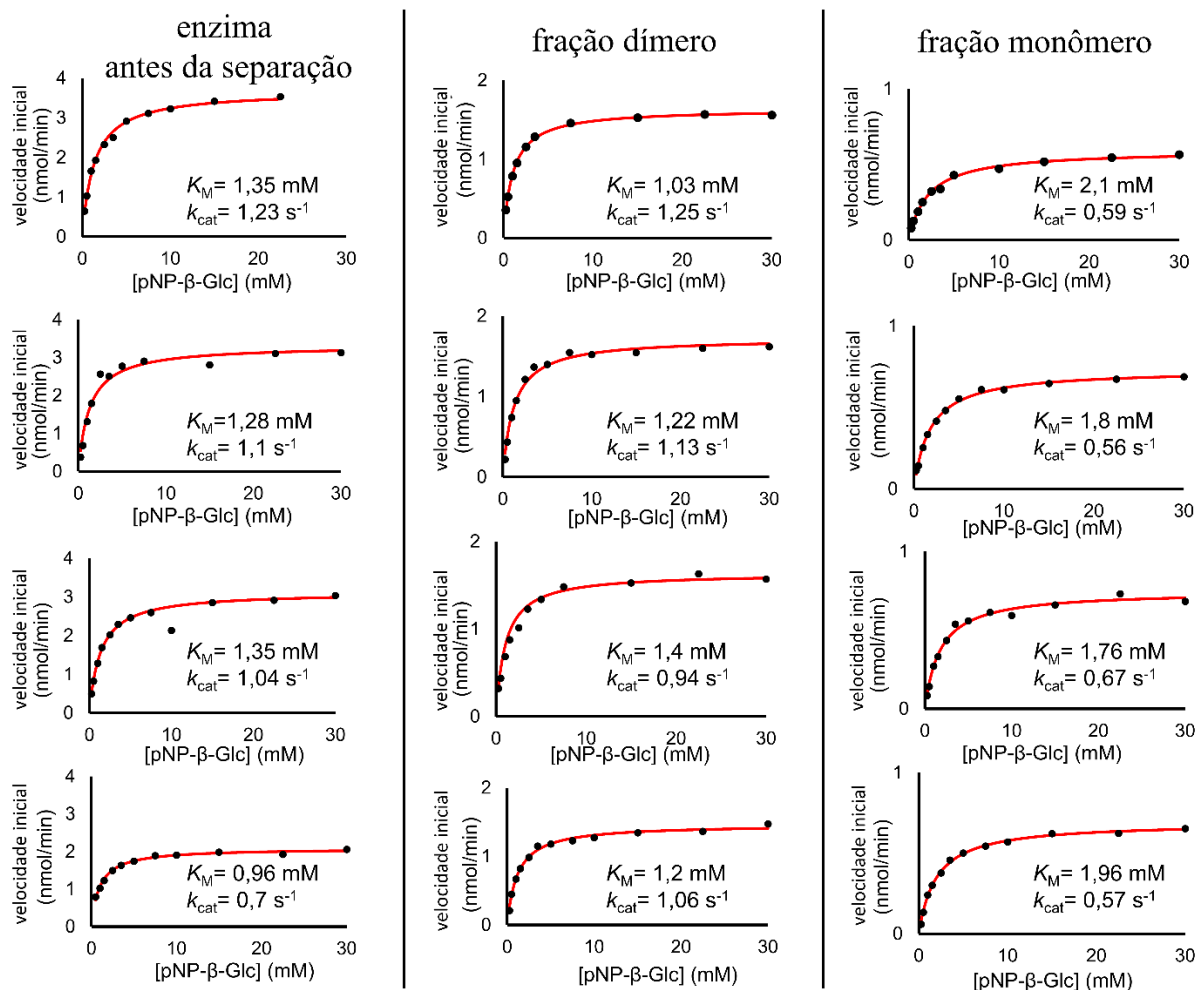


Figura 1.13. Determinação de parâmetros de cinética enzimática da Sfβgly purificada antes do SEC e nas frações posteriores à separação nesta cromatografia. Quatro experimentos independentes foram realizados para cada grupo. Linha em vermelha indica o ajuste do modelo de Michaelis–Menten aos dados experimentais. Todos os experimentos foram feitos em tampão F100.

Vale ressaltar que nos experimentos acima, o equilíbrio $M + M \rightleftharpoons D$ seria desviado para a esquerda, já que a concentração da enzima nos ensaios ($0,5 \mu\text{M}$) é cerca de 10 vezes abaixo do K_D ($3,7 \mu\text{M}$). Então, a eficiência catalítica da fração dimérica pode estar subestimada se uma quantidade de enzima se dissocia para a forma monomérica significativamente durante o tempo de ensaio (20 min). Contudo, vale notar que os k_{cat} e K_M de dímeros e monômeros são diferentes da média ponderada esperada na amostra de enzima antes da separação por SEC. Esta observação sugere que embora possa haver conversão de um estado oligomérico em outro, este processo ainda não atingiu o equilíbrio. Deste modo, os parâmetros cinéticos são representativos de dímeros e monômeros isolados.

Diante da evidência de que a homodimerização influencia a atividade da Sfβgly, procurou-se analisar a cinética da dissociação do dímero, que como será visto na próxima sessão, gerou e sustentou uma hipótese sobre o mecanismo de dimerização, o qual unificou a oligomerização e seu efeito na atividade da Sfβgly.

1.3.3 – Cinética de homodimerização da Sfβgly e suas conformações

Com os resultados dos experimentos anteriores, reunimos indícios que o processo de homodimerização é descrito por um equilíbrio, porém não foi viável detalhar seu mecanismo. Portanto, decidiu-se investigar a oligomerização através de experimentos que acompanhem este processo de forma contínua. Assim, foram feitos experimentos para análise da fluorescência intrínseca dos resíduos de triptofano de Sfβgly. Testes com excitação a $EX_{\lambda} = 295$ nm em diferentes concentrações de Sfβgly purificada mostraram espectros de emissão com pico em $EM_{\lambda} = 337$ nm (**Figura 1.14a**), que são característicos para proteínas enoveladas (Shukla & Guptasarma, 2004). A intensidade de fluorescência aumenta quase linearmente de 0,5 até 15 μ M de Sfβgly, mas curiosamente diminui conforme aumenta a concentração na solução acima de 15 μ M. Esse efeito da concentração ficou mais claro quando as intensidades de fluorescência em 337 nm, agora chamadas de F_{337} , foram plotadas com respeito a [Sfβgly] (**Figura 1.14b**).

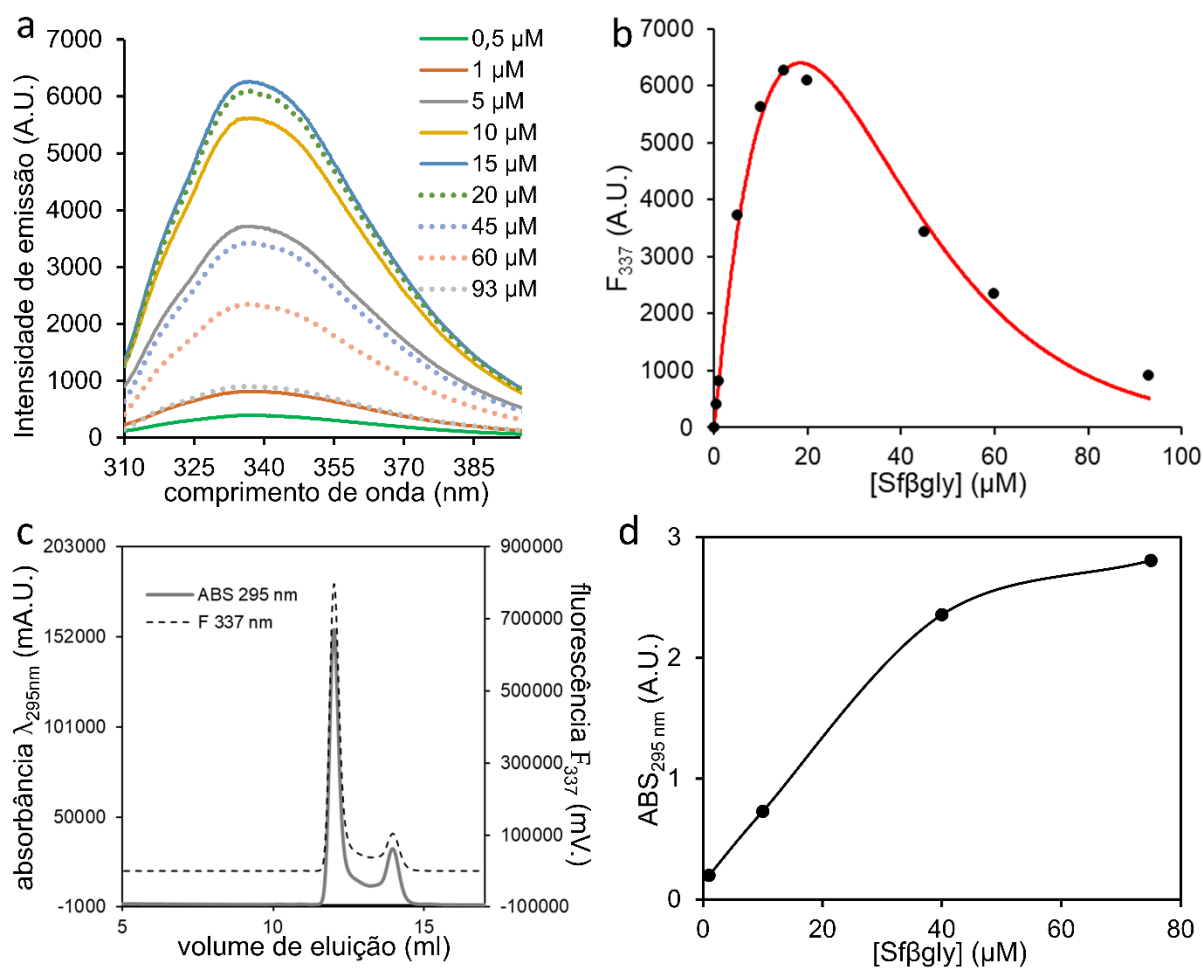


Figura 1.14. Fluorescência intrínseca e absorbância da Sfβgly. **a:** Espectro da fluorescência ($EX_{\lambda} = 295\text{ nm}$) de Sfβgly em diferentes concentrações em tampão F100; **b:** intensidade de fluorescência medida no $EM_{\lambda} = 337\text{ nm}$ para cada [Sfβgly] testada; **c:** fluorescência ($EX_{\lambda} = 295\text{ nm}$, $EM_{\lambda} = 337\text{ nm}$) e absorbância (295 nm) dos picos de dímero e monômero de amostra de Sfβgly em sistema HPLC–SEC com coluna Superdex™ 75 10/300 (GE Healthcare, SE); **d:** medidas de absorbância a 295 nm de amostras de Sfβgly.

Inicialmente considerou-se que a explicação para este fenômeno de “apagamento” da fluorescência conforme se aumenta a concentração de enzima (Figura 1.14b) poderia estar relacionada com o acúmulo do dímero em solução, já que em maiores concentrações da enzima mais dímeros são observados (Figura 1.11). Mas, mostramos experimentalmente que a fluorescência intrínseca correspondente aos dímeros de Sfβgly, é cerca de 1,5 vezes maior que a do pico que corresponde ao monômero (Figura 1.14c). Deste modo, a diminuição da fluorescência intrínseca nas maiores [Sfβgly] não poderia ser devida ao acúmulo de dímeros. De fato, a diminuição de fluorescência em altas concentrações de Sfβgly deve estar relacionada com um efeito de filtro interno, pois a absorbância

em 295 nm das amostras de enzima decai acima de certa concentração (**Figura 1.14d**), evidenciando que com o aumento de concentração apenas parte da amostra é efetivamente excitada em 295 nm.

Deste modo, o gráfico na **Figura 1.14b** pode ser descrito como $F_{337} = (F_{337D} \cdot 2[D] + F_{337M} \cdot [M]) \cdot e^{-Q[Sfbgly]}$, sendo F_{337D} e F_{337M} representam a fluorescência intrínseca específica do dímero e do monômero, respectivamente. “ Q ” (de “*quenching factor*”) corresponde a um fator de apagamento da fluorescência, cuja magnitude depende da concentração de Sfbgly. As $[D]$ e $[M]$ foram calculadas a partir da $[Sfbgly]$ levando em consideração as frações de dímero já estimadas na **Figura 1.11**. O ajuste desta equação nos dados experimentais (**Figura 1.14b**; linha vermelha) resultou em uma fluorescência intrínseca do dímero (F_{337D}) 1,9 vezes maior que aquela do monômero, valor próximo com a medida experimental (1,5 vezes na **Figura 1.14c**). Com isto, ficou demonstrado que a intensidade da fluorescência intrínseca (F_{337}) depende da abundância relativa dos dímeros e monômeros de Sfbgly, sendo que os dímeros são mais fluorescentes do que os monômeros, provavelmente devido a dois triptofanos, W163 e W305, que se encontram escondidos do solvente na interface de dimerização (**Figura 1.15**).

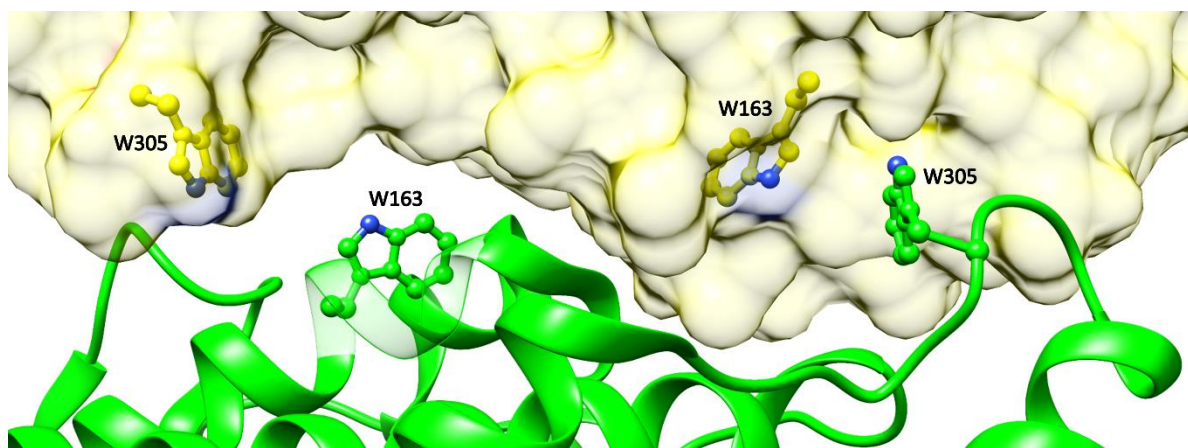


Figura 1.15. Resíduos W163 e W305 na interface do homodímero. Subunidades de Sfbgly estão distinguidas por cores e forma de representação da cadeia.

Considerando que a fluorescência intrínseca de Sfbgly (F_{337}) é relacionada com a abundância relativa de dímeros e monômeros, foi planejado uma estratégia experimental para mensurar a cinética de conversão entre estes estados. Uma solução de alta concentração da Sfbgly tem o seu equilíbrio de estados perturbado quando diluído para uma concentração abaixo do K_D e acompanhando a variação

de F_{337} , seria possível verificar a dissociação do dímero e formação de monômero ao longo do tempo. Nas **Figuras 1.16a,b** são mostrados o desenvolvimento de F_{337} quando uma rápida diluição na amostra é feita de $93 \mu\text{M}$ para $0,5 \mu\text{M}$ em função do tempo. Deve-se atentar ao fato de que a amostra é diluída e somente então se inicia o registro da fluorescência intrínseca, assim o tempo inicial já corresponde à leitura após a diluição. Nota-se que F_{337} diminui em uma escala de tempo de dezenas a centenas de minutos, significando que monômeros livres, que possuem menor fluorescência intrínseca, estão sendo formados. A curva de dissociação do dímero é bem descrita com um modelo de exponencial simples ($F_t = Ae^{-k_{obs}t} + F_{t_0}$, linha vermelha **Figura 1.16a**) gerando, uma constante de decaimento (k_{obs}), pela qual o sistema atinge o equilíbrio, “A” referente a amplitude do sinal, e “ F_{t_0} ” que é a fluorescência inicial.

Também foi possível atestar a dissociação do dímero através da sua cinética enzimática (**Figura 1.16b**), que demonstraram um decaimento de k_{cat}/K_M similar com a redução de fluorescência ($R^2 = 0,966$), ou seja, com a diminuição na concentração de dímeros em solução. Deve ser notado que todas as amostras de enzima usadas para determinação de k_{cat}/K_M estavam na mesma diluição, ou seja, continham a mesma concentração de enzima total.

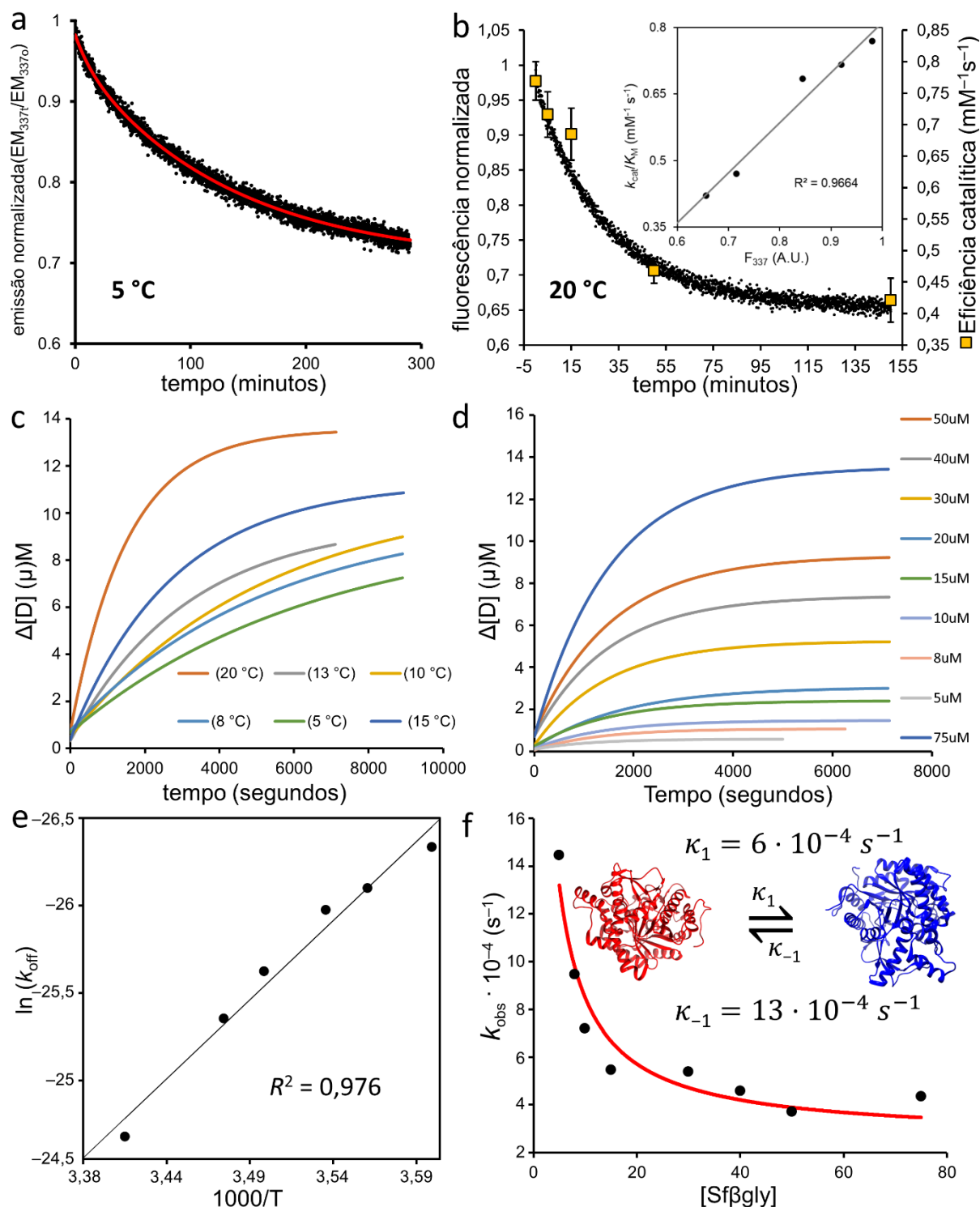


Figura 1.16. Cinética de dissociação do homodímero de Sfβgly. **a**: F_{337} de Sfβgly após diluição de 93 μM para 0,5 μM a 5 °C; **b**: F_{337} e eficiência catalítica de Sfβgly após diluição a 20 °C e gráfico de correlação entre as variáveis F_{337} e k_{cat}/K_M ($R^2 = 0,966$); **c,d**: variação da concentração de dímero após diluição de Sfβgly após diluição de 93 μM para 0,5 μM em diferentes temperaturas e de diferentes $[\text{Sf}\beta\text{gly}]_0$ antes da diluição para 0,5 μM ; **e**: gráfico de Eyring correlacionando o logaritmo natural da constante de formação de monômero (k_{off}) com o inverso da temperatura absoluta. k_{off} foi estimado a partir dos primeiros 3 min após a diluição de Sfβgly; **f**: efeito da $[\text{Sf}\beta\text{gly}]_0$ na constante de formação de monômero

(k_{obs}). O ajuste das medidas foi feito com uma equação modificada do trabalho de Vogt A.D. & Di Cera E. (Vogt & di Cera, 2012) considerando um mecanismo de seleção de confôrmeros descrito na “Equação 1” para determinar as constantes cinéticas de troca conformacional k_1 e k_{-1} .

Com o método de diluição rápida e fluorescência intrínseca, foram feitos vários ensaios variando-se a temperatura e concentração inicial da enzima (**Figura 1.16c,d**), os quais foram analisados considerando a variação da concentração de dímero ($\Delta[D]$) calculada empregando-se a relação $\Delta[D] = ((F_{337t}/F_{337t0} \cdot [Sfbgly]) - [Sfbgly])/2$.

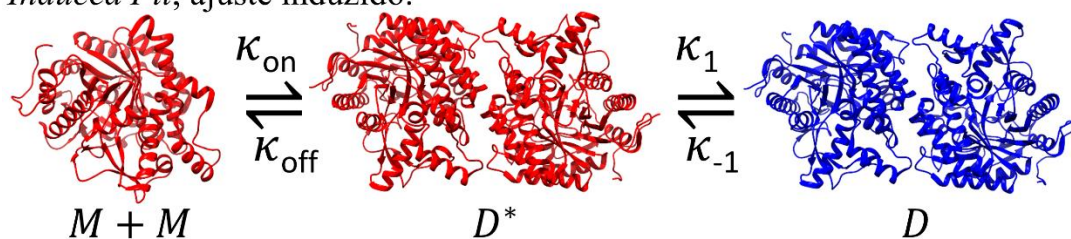
Usando os dados dos experimentos executados em diferentes temperaturas foi possível obter a constante de dissociação do dímero, k_{off} , a partir da velocidade inicial de cada curva (primeiros 3 min), intervalo no qual considerou-se que ocorreria essencialmente apenas a seguinte reação: $D \rightarrow 2M$. Com os k_{off} obtidos foi elaborado o gráfico de Eyring (**Figura 1.16c,e**), a partir do qual obtivemos a energia de ativação do processo de dissociação do dímero, $E_a = 18,5 \text{ kcal mol}^{-1}$.

A seguir os dados de $\Delta[D]$ e tempo obtidos nos experimentos variando a concentração inicial de enzima foram ajustados na equação exponencial $\Delta[D] = Ae^{k_{obs}t} + \Delta[D]_0$, rendendo o k_{obs} , conforme já comentado acima. Para entender os k_{obs} , mostrados na **Figura 1.16d,f**, conjecturou-se que o processo para alcançar o novo equilíbrio entre dímeros e monômeros após a perturbação (diluição) poderia ser descrito por mais de uma etapa.

Dois modelos de homodimerização para o caso de Sfbgly são considerados e mostrados na **Figura 1.17**, ajuste induzido e seleção de confôrmero (Csermely et al., 2010). De fato, estes dois modelos são casos extremos em um gradiente de mecanismos, onde inclusive há combinações entre ambos. Mas, de qualquer modo, são úteis para uma abordagem simplificada deste problema. Como já estabelecido na literatura, assumindo *a priori* que as trocas conformacionais são mais lentas do que os eventos de ligação, condição de equilíbrio rápido, no caso do mecanismo do encaixe induzido as constantes de velocidade (k_{obs}) pelo qual o sistema evolui para o equilíbrio se relaciona com a concentração de proteína–ligante de modo ascendente até um patamar. Já no caso do mecanismo da seleção de confôrmero, a relação é inversa, partindo de um k_{obs} mais alto e decaindo até um patamar inferior (Vogt & di Cera, 2012). Assim, analisando graficamente os k_{obs} extraídos da **Figura**

1.16d, nota-se que o perfil encontrado na **Figura 1.16f** enquadra-se no modelo de seleção de confôrmero.

Induced Fit, ajuste induzido.



Conformational Selection, seleção de confôrmeros

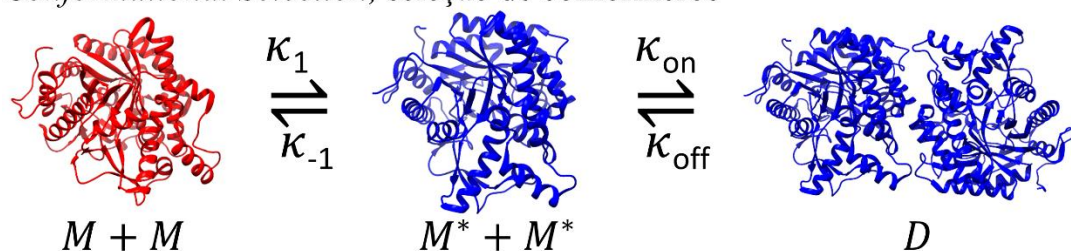


Figura 1.17. Modelos de homodimerização de Sfbgly com mudança conformacional. k_{on} e k_{off} referem-se respectivamente ao evento de ligação e dissociação das subunidades, k_1 e k_{-1} referem-se ao evento de mudança conformacional. M e D indicam os estados monomérico e dimérico, respectivamente. As cores vermelha e azul indicam distintas conformações. * indica um intermediário com conformação diferente.

Contudo, para uma análise quantitativa mais detalhada é importante discutirmos alguns aspectos. Além da condição de equilíbrio rápido, os modelos apresentados acima e o comportamento de k_{obs} são analisados em uma condição experimental chamada de “pseudo-primeira ordem”, em que a concentração do ligante é muito maior que da proteína “ligadora”, e deste modo, não se altera ao longo do processo até o equilíbrio. No caso do processo de homodimerização de Sfbgly não é possível atender a condição de pseudo-primeira ordem, uma vez que a proteína é o ligador e o ligante simultaneamente, assim a formação do dímero obrigatoriamente altera a concentração do ligante. Esta é uma limitação que certamente dificulta a análise de qualquer processo de homodimerização.

Para contornar esta dificuldade, tomamos da literatura a informação de que a constante de velocidade de processos de ligação entre monômeros (k_{on}) pode ser aproximada para o formato de pseudo-primeira ordem quando expressa como $4k_{\text{on}}[M]$ (Gutfreund, 1995). Assim, usando esta formatação introduzimos esta constante de velocidade na equação geral para o k_{obs} esperado nos mecanismos de seleção de confôrmero e encaixe induzido, sendo que as expressões simplificadas

obtidas, embora agora pontualmente diferentes, ainda mantêm o mesmo perfil geral em função da concentração de proteína total. Especificamente no caso do mecanismo de seleção de confôrmeros estabelecemos a seguinte expressão para k_{obs} :

Equação 1:
$$k_{obs} = k_1 + k_{-1} \left(\frac{1}{1 + 4 \frac{k_{on}}{k_{off}} [M]_{total}} \right)$$

Assim, ajustando os dados experimentais da **Figura 1.16f** na equação acima foi possível determinar k_1 e k_{-1} do processo de troca da conformação dos monômeros de Sfβgly, o qual ocorre antes do evento de ligação. Com o K_D obtido nos experimentos de SEC (**Figura 1.11**) calculou-se a variação de energia livre da dissociação do homodímero de Sfβgly, $\Delta G = 6,8 \text{ kcal mol}^{-1}$, enquanto que a partir dos experimentos de rápida diluição em diversas temperaturas estimamos a energia de ativação ($E_a = 18,5 \text{ kcal/mol}$) (**Figura 1.16e**) do processo de dissociação do dímero de Sfβgly. A constante de dissociação foi estimada diretamente da velocidade inicial dos ensaios de diluição rápida ($k_{off} = 61 \cdot 10^{-6} \text{ s}^{-1}$ a $5 \text{ }^\circ\text{C}$) e a constante de associação (k_{on}) foi estimada para satisfazer a relação $K_d = \frac{k_{off}}{k_{on}}$, ou seja, $k_{on} = 16,5 \text{ M}^{-1}\text{s}^{-1}$.

Na **Figura 1.18** é mostrado o modelo que resume aspectos termodinâmicos e cinéticos da homodimerização da Sfβgly através do modelo de seleção de confôrmeros. Desta forma, a subunidade monomérica visita um estado conformacional de baixa afinidade para formar o dímero, a qual exibe menor k_{cat}/K_M . Mas, uma troca de conformação ocorre de forma lenta (escala de minutos), levando a uma forma monomérica apta à dimerização, a qual exibe maior k_{cat}/K_M . Para capturar e estabilizar este novo estado de maior k_{cat}/K_M , é preciso o encontro com outro monômero, também de alta afinidade, e assim estabiliza-se gerando o homodímero mais ativo. Notavelmente, propomos que a troca de conformações é acompanhada por mudança na atividade catalítica da Sfβgly. Em suma, unificamos a ocorrência dos diferentes estados oligoméricos e de suas respectivas atividades catalíticas através da troca conformacional que ocorreria em Sfβgly.

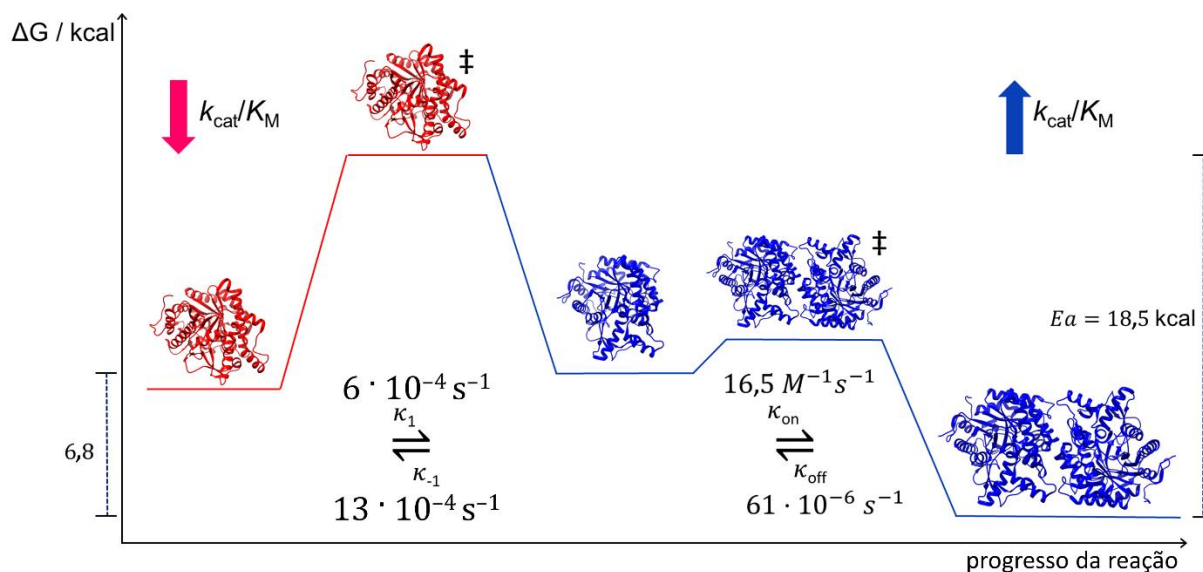


Figura 1.18. Modelo termodinâmico e cinético da homodimerização de Sfbgly por seleção de conformêro. O estado monomérico em vermelho é também o de menor eficiência catalítica, que sofre uma mudança conformacional para um estado em azul que é capaz de se dimerizar e mais ativo. Os termos energéticos e cinéticos foram estimados com ensaios feitos a 5 °C. Figura originalmente de Otsuka F.A.M. et al. (Otsuka et al., 2020).

1.3.4 – Conexão da interface para o sítio ativo

As evidências experimentais deixaram claro que as diferentes eficiências catalíticas e distintos estados de oligomerização estão relacionados, mas a explicação molecular deste fenômeno ainda não é clara, em especial por conta da ausência de modelo estrutural do estado monomérico. Felizmente, ano passado (2021) a DeepMind disponibilizou o AlphaFold2 (Baek et al., 2021; Jumper et al., 2021), uma ferramenta computacional que provocou uma revolução em biologia estrutural, pois permite a predição da estrutura tridimensional de qualquer proteína a partir de sua sequência utilizando redes neurais com alta confiabilidade e precisão que rivaliza com métodos experimentais. Também foi reportado que a predição do AlphaFold2 é capaz de capturar estados conformacionais se indicado o estado oligomérico (Cummins et al., 2022).

A estrutura da Sfbgly monomérica predita por AlphaFold2 e disponível no banco de dados da Uniprot (id: O61594) foi comparada com a estrutura cristalográfica do dímero (Tamaki et al., 2016) na **Figura 1.19a**. A comparação estrutural mostrou que a estrutura global é quase idêntica, com um RMSD = 0,26 Å entre todos os pares de resíduos, assim como os resíduos que fazem contato com

o substrato no sítio ativo. Para avaliar com mais detalhe as possíveis diferenças conformacionais entre as estruturas, foi usada a teoria de redes na análise dos contatos entre os pares de resíduos (Brinda & Vishveshwara, 2005). Desta forma, cada resíduo da proteína é representado como um nó de uma rede formada por vértices que representam contatos não-covalentes com outros resíduos a menos de 5 Å de distância. Para analisar os resíduos individualmente, utilizamos a pontuação de centralidade de intermediação (*betweenness*) de caminhos mais curtos, que atribui a cada resíduo uma medida de quão central ele é, *i.e.*, quantifica o quanto dado resíduo ou nó age como um conector ao longo do caminho mais curto entre outros nós.

Na **Figura 1.19b**, analisando as redes interação de resíduos (RIN, do inglês *Residue Interaction Network*) das duas estruturas, nota-se que a maioria dos resíduos permanece com a sua centralidade quase inalterada considerando um limiar de 0,5 na variação, mas os resíduos Y141, W143, N186 e T245 se destacaram dentre aqueles que mais sofrem variação em seu índice de centralidade de intermediação. A diferença na centralidade destes resíduos não parece ser causado por mudanças de contatos dos seus vizinhos mais imediatos, como mostrados na **Figura 1.19c,d**.

Verificando o posicionamento estrutural dos resíduos Y141, W143, N186 e T245 (**Figura 1.20**), nota-se que eles formam uma ponte de contato direto entre a α -hélice da interface com sítio ativo, ou seja, podem ser os responsáveis por transmitir os ajustes conformacionais dos resíduos na interface para o sítio ativo quando ocorre a mudança conformacional nos monômeros que favorece a dimerização e o aumento de atividade catalítica.

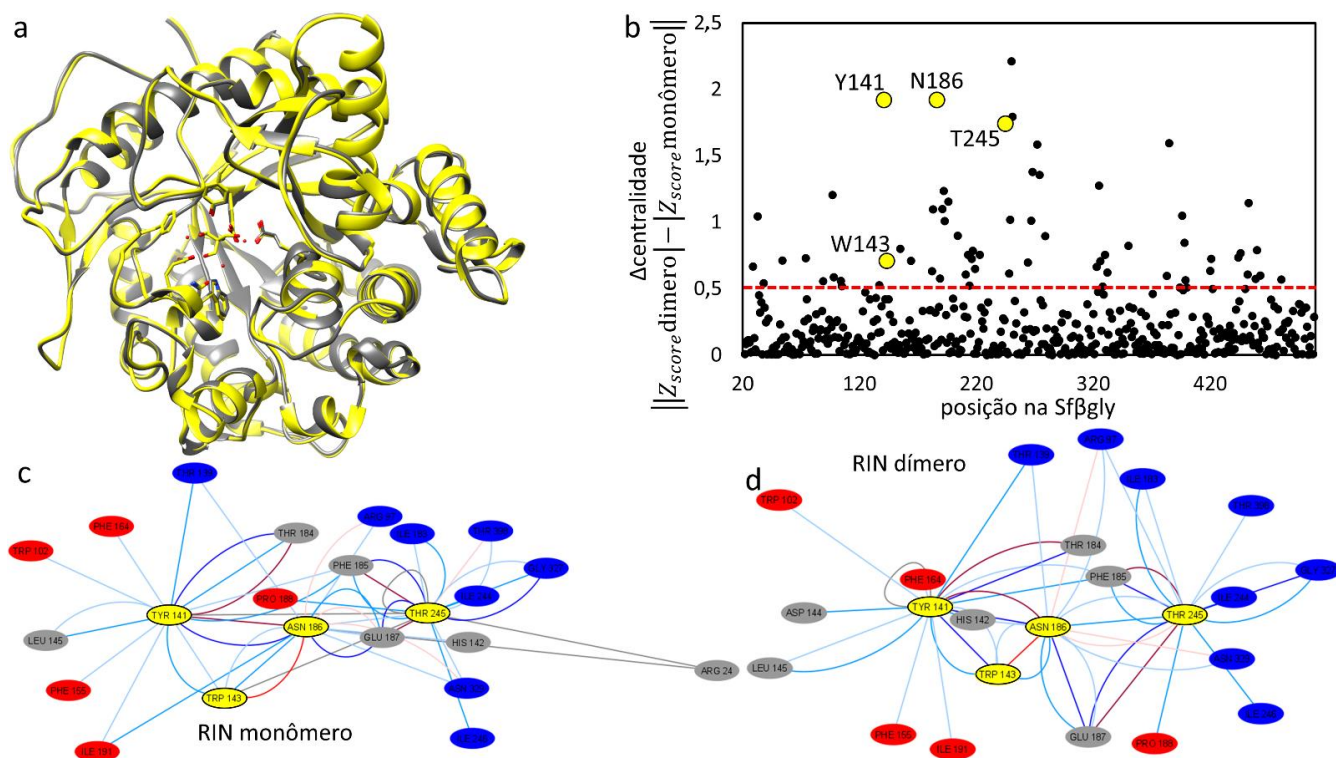


Figura 1.19. a: Comparação das estruturas de Sfβgly determinadas por cristalografia em amarelo (*i.e.*, dímero) e predita computacionalmente por AlphaFold2 em cinza (*i.e.*, monômero). b: Diferença de centralidade *betweenness* entre pares de resíduos das duas estruturas, destacando em amarelo os resíduos que conectam a interface com o sítio ativo, e um limiar de relevância de 0,5 na diferença de pontuação em linha tracejada vermelha. c, d: Rede de interações de resíduos (RIN) entre os quatro resíduos (em amarelo) selecionados Y141, W143, N186 e T245; cores vermelho, cinza e azul dos resíduos indicam local de estrutura secundária em α -hélice, desordenado e folha- β , respectivamente. As RIN foram construídas usando o *software* Cytoscape.

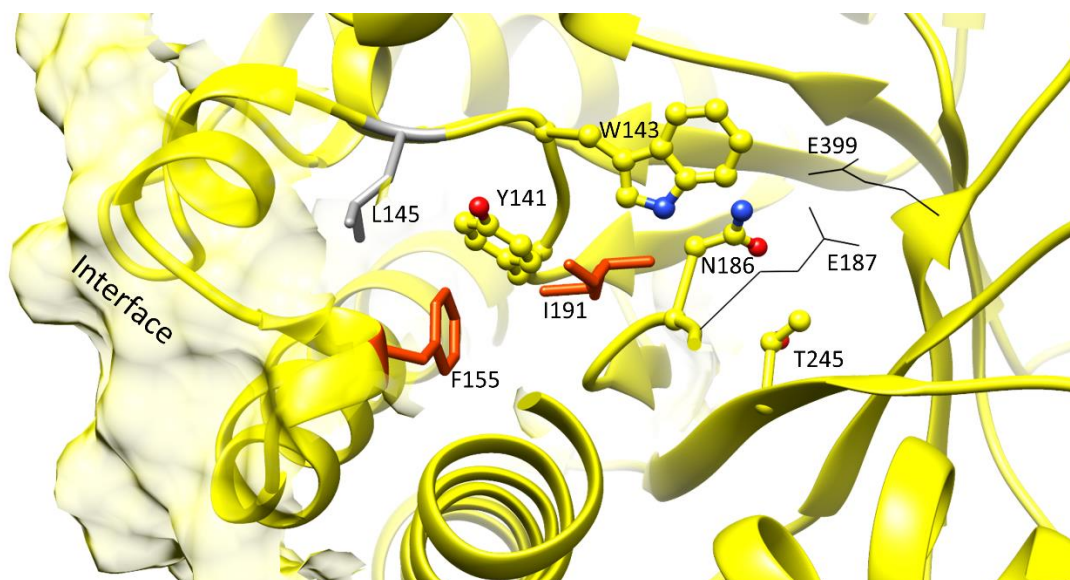


Figura 1.20. Rede de resíduos entre a interface (em superfície transparente) e o sítio ativo (delimitado pelos resíduos W143, N186, E399 e E187) na estrutura cristalográfica de

Sfbgly. As cadeias laterais mostradas dos resíduos Y141, W143, N186 e T245 são os que tiveram suas centralidades fortemente alteradas (**Figura 1.19b**). Resíduos em vermelho e cinza indicam local de estrutura secundária em α -hélice e *loop*, respectivamente.

Tendo em mente esta hipótese, para especular sobre a dinâmica dos ajustes conformacionais dos resíduos tanto na interface quanto no sítio ativo, usamos a ferramenta *Morpho Conformations*, disponível no *software* de visualização molecular UCSF Chimera (Pettersen et al., 2004), para simular as transições conformacionais de dímero para monômero e *vice-versa* (**Figura 1.21**).

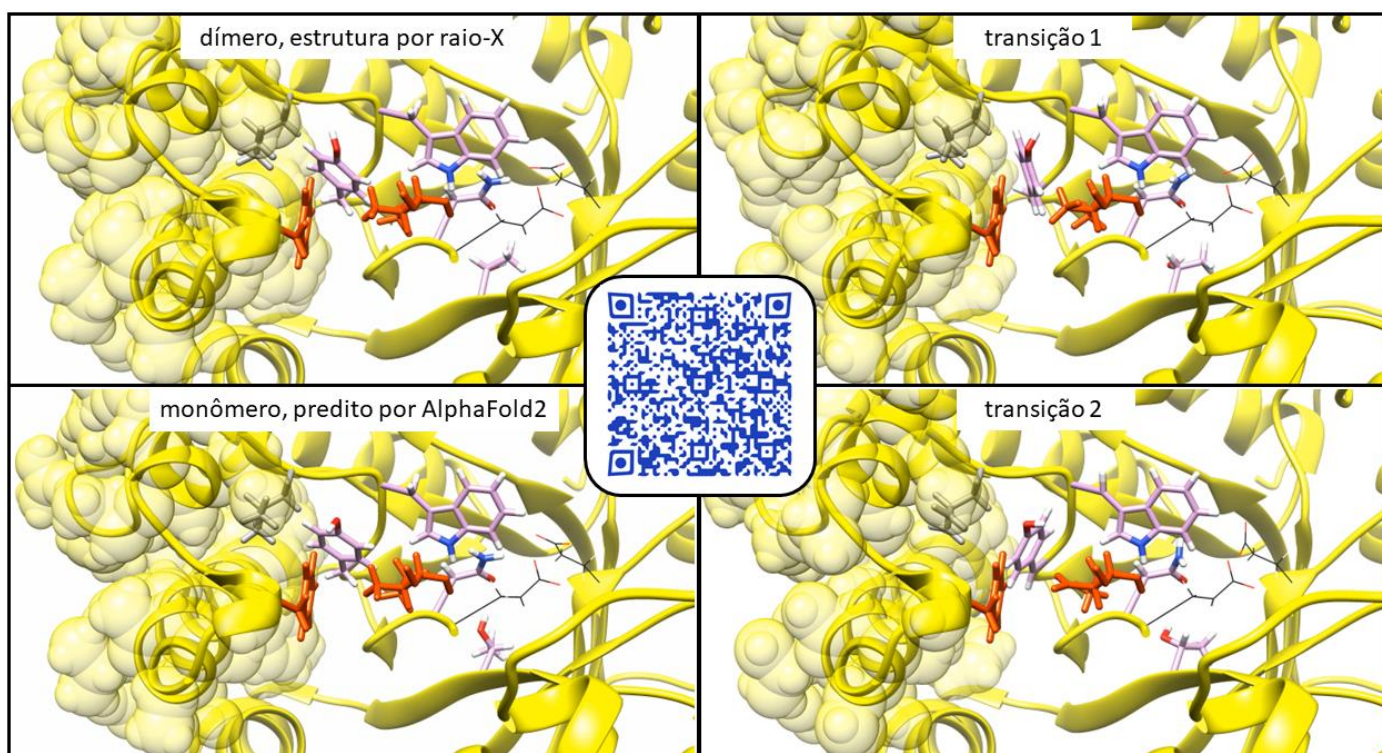


Figura 1.21. Animação ilustrando a mudança conformacional entre dímero e monômero de Sfbgly, destacando em lilás os resíduos que tiveram seus *Z-score* de centralidade *betweenness* alterados. O QR-code direciona para a animação em formato .gif com duração de 3 s, e mostra as transições na seguinte ordem: dímero \rightarrow monômero \rightarrow dímero. Sequência de *frames* feito com a ferramenta *Morpho Conformations* do UCSF Chimera (Pettersen et al., 2004), arquivo animado .gif feito em <https://giphy.com/>, QR-code criado em <https://www.qrcode-monkey.com/>. A animação também está disponível no link abaixo: (<https://media2.giphy.com/media/k9Bad4dN9Kakqmgf9w/giphy.gif?cid=790b7611105b42203fefcb84e150872ff248dab9a0f25741&rid=giphy.gif&ct=g>)

Na animação, disponível pelo QR-code ou *link* indicado na legenda, são mostradas grandes movimentações nos resíduos que compõem a interface do dímero, em conjunto com o resíduo Y141 que gira quase 180° em torno de seu eixo devido a um espaço adicional gerado logo abaixo do anel aromático de F155.

Também são vistas grandes mudanças nos rotâmeros dos resíduos I191 e T245, que estão em contato com resíduos no sítio ativo responsáveis pela catálise. Em suma, os modelos estruturais de Sfbgly, cristalográfico e predito pelo AlphaFold, oferecem uma visão molecular da transição de conformação associada à dimerização, mostrando os resíduos críticos que conectam a interface com o sítio ativo, e embasando a hipótese de que uma alteração conformacional que favorece a dimerização também resulta em aumento da atividade catalítica desta enzima.

1.3.5 – Mutantes de Sfbgly e suas implicações na oligomerização e atividade enzimática

Após entender o mecanismo do processo de oligomerização e sua importância para a atividade da enzima Sfbgly, optou-se por verificar se mutações na interface do dímero modulariam a sua estabilidade. As seguintes abordagens foram praticadas: primeiramente foram investigadas mutações simples que desestabilizariam a formação do homodímero. Na segunda, usamos métodos computacionais que utilizam algoritmos especializados em otimizar as energias de interação na interface do dímero através de mutações a partir de uma estrutura cristalográfica oferecida como *input*. O motivo para obtenção de um dímero da Sfbgly mais estável seria ter uma enzima mutante estabilizada no seu estado de maior atividade catalítica através de mutações distantes do sítio ativo, portanto, com menos impacto em sua função. Porém, adiantamos que as mutantes planejadas para estabilizar o homodímero não eram solúveis o suficiente para serem purificadas e testadas. Portanto, até o momento não há uma forma estritamente homodimérica da Sfbgly para caracterização.

Por outro lado, foi possível analisar sete mutantes pontuais de resíduos que estão na interface do dímero que o desestabilizaram. Três destas mutações na interface (**Figura 1.22a**) foram a troca de um resíduo hidrofóbico para um polar (L152K, A207K e M110E), o que permitiu avaliar se a hidrofobicidade da interface é um elemento crítico para a dimerização. Os três mutantes resultaram na desestabilização completa do homodímero segundo evidenciado em ensaios de SEC (**Figura 1.23**). Confirmando que o efeito hidrofóbico (Southall et al., 2002) é um importante fator para a tornar o dímero termodinamicamente favorável. Por outro

lado, as trocas de resíduos polares de cadeia lateral longa (**Figura 1.22b**) para resíduo polar de menor cadeia polar (N157S e D166S) resultaram em uma leve desestabilização do dímero (**Figura 1.23**). Portanto, as interações polares formadas por N157 e D166 na interface, embora também contribuam para estabilidade do dímero, não são o principal fator de estabilização em termos energéticos, mas talvez contribuam com a “especificidade” de dimerização.

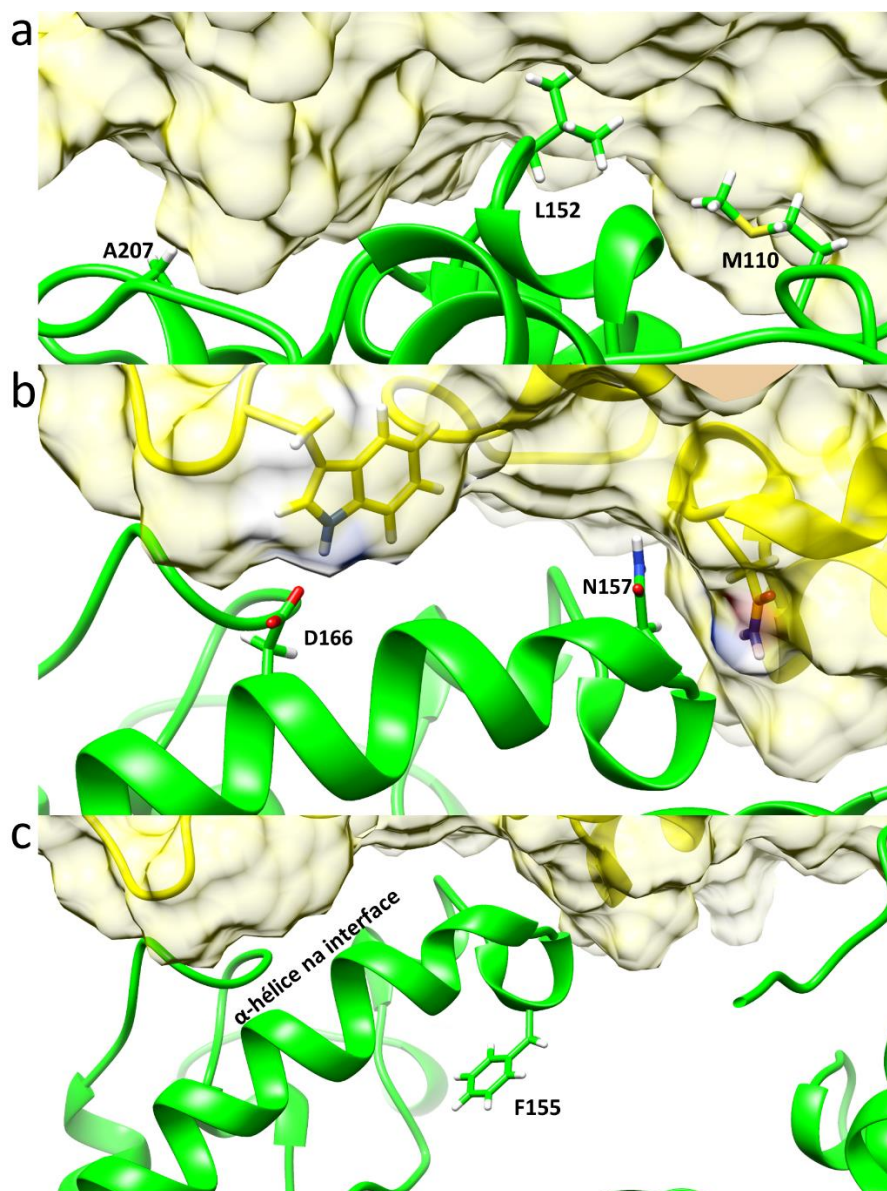


Figura 1.22. Posição dos resíduos na interface do homodímero escolhidos racionalmente para mutação simples. **a:** resíduos hidrofóbicos M110, L152 e A207; **b:** resíduos polares N157 e D166; **c:** Resíduo F155, que estabiliza uma α -hélice da interface.

Finalmente foram analisados dois mutantes, F155A e F155N (**Figura 1.22c**), que em tese modificariam a estabilidade de uma α -hélice que faz muitos contatos com a interface. A mutação F155A tornou a Sfβgly predominantemente

monomérica, sugerindo que o contato da α -hélice com a interface é importante para a estabilização do dímero. Por outro lado, F155N prejudicou a solubilidade e purificação da enzima, portanto, é sugerido que estruturas secundárias na interface do dímero são um importante fator para a tornar a dimerização termodinamicamente favorável.

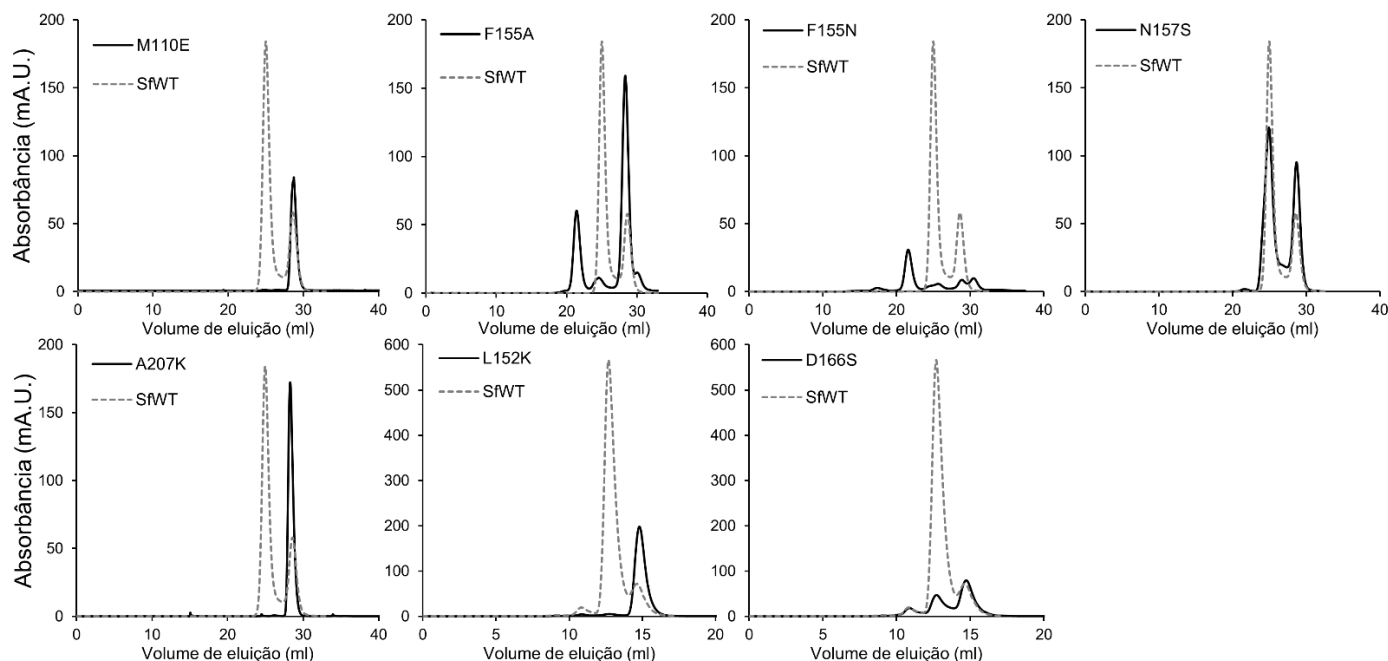


Figura 1.23. SEC de mutantes pontuais de Sfbgly. Cromatograma da enzima selvagem (20 μ M) em pontilhado é comparado com as mutantes em linha sólida M110E (15 μ M), F155A (20 μ M), F155N (3 μ M), N157S (27,5 μ M), A207K (50 μ M), L152K (24 μ M) D166S (19 μ M). Mutantes M110E, F155A, F155N, N157S e A207K foram analisados com duas colunas Superdex 200 acopladas e mutantes L152K e D166S foram analisados com uma coluna Superdex 200 Increase. Todas as cromatografias foram feitas a 5 °C.

Com relação às atividades enzimáticas das mutantes pontuais, a **Tabela 1.3** revela que as mutantes M110E, L152K, A207K e F155A tiveram suas eficiências catalíticas, expressas como k_{cat}/K_M , reduzidas na amostra antes do SEC, porém parecidas com a da fração monomérica da enzima selvagem após a separação em SEC, concordando com a desestabilização do dímero. Mas, as mutantes N157S e D166S, têm sua eficiência catalítica similar à da selvagem na amostra antes do SEC, coerente com trocas que brandamente desestabilizam o dímero. De modo geral, as mutações simples testadas na interface impactaram negativamente na dimerização, conseqüentemente, preveniram o acesso da conformação de maior eficiência catalítica mostrado anteriormente.

Por fim, o mutante planejado para manter o dímero estável através de uma ligação dissulfeto entre suas subunidades através da troca N157C (**Figura 1.22b**), não foi obtido mesmo depois de várias tentativas de mutagêneses.

Tabela 1.3. Parâmetros cinéticos de mutantes pontuais da Sfbgly purificadas empregando o substrato *p*-nitrofenil β -glucosídeo antes da separação por SEC e das frações posteriores à separação. O símbolo “–” indica que não houve atividade detectável. Dados são médias \pm SEM.

	Enzima antes da SEC			Enzima após SEC	
	K_M (mM)	k_{cat} (s ⁻¹)	k_{cat}/K_M (mM ⁻¹ s ⁻¹)	dímero k_{cat}/K_M	monômero k_{cat}/K_M
Sfbgly WT	1,3 \pm 0,1	1,23 \pm 0,01	0,9 \pm 0,1	1,03 \pm 0,04	0,32 \pm 0,04
M110E	2,9 \pm 0,3	0,26 \pm 0,01	0,09 \pm 0,01	–	0,10 \pm 0,02
L152K	2,3 \pm 0,7	0,7 \pm 0,1	0,3 \pm 0,1	–	0,4 \pm 0,1
A207K	1,7 \pm 0,1	0,9 \pm 0,1	0,5 \pm 0,1	–	0,25 \pm 0,05
F155A	1,8 \pm 0,1	0,65 \pm 0,02	0,35 \pm 0,03	–	–
F155N	–	–	–	–	–
N157S	2,7 \pm 0,7	1,2 \pm 1,1	0,4 \pm 0,2	1,0 \pm 0,1	0,13 \pm 0,04
D166S	1,2 \pm 2,0	1,0 \pm 0,1	0,83 \pm 0,05	1,0 \pm 0,3	0,7 \pm 0,6

1.3.6 – *Design* de variante de Sfbgly com capacidade de hidrolisar α - e β -substrato

Na última década, testemunhamos avanços biotecnológicos singulares na predição da estrutura de proteínas e modificação da sua função sob medida (Kuhlman & Bradley, 2019). Através de *softwares* como o Rosetta (Leaver-Fay et al., 2011), é possível modelar e obter sequências de aminoácidos que irão possuir estruturas tridimensionais e funcionais (Bjelic et al., 2013; Dou et al., 2018). Assim, submetemos a Sfbgly em dois programas de *design* de sequência, primeiro no programa PROSS (Goldenzweig et al., 2016) e depois no FuncLib (Khersonsky et al., 2018), que utilizam o Rosetta para gerar variantes de maior estabilidade e de maior repertório enzimático, *i.e.*, permitir novos substratos e modular a eficiência catalítica dos substratos já conhecidos.

Na **Figura 1.24** é mostrado a variante da Sfbgly que acumula 43 mutações após o uso do *software* PROSS, somadas com mais 4 mutações (Y38W, W143F,

L372M e S445C) feitas pelo FuncLib, totalizando 47 mutações em comparação com a selvagem.

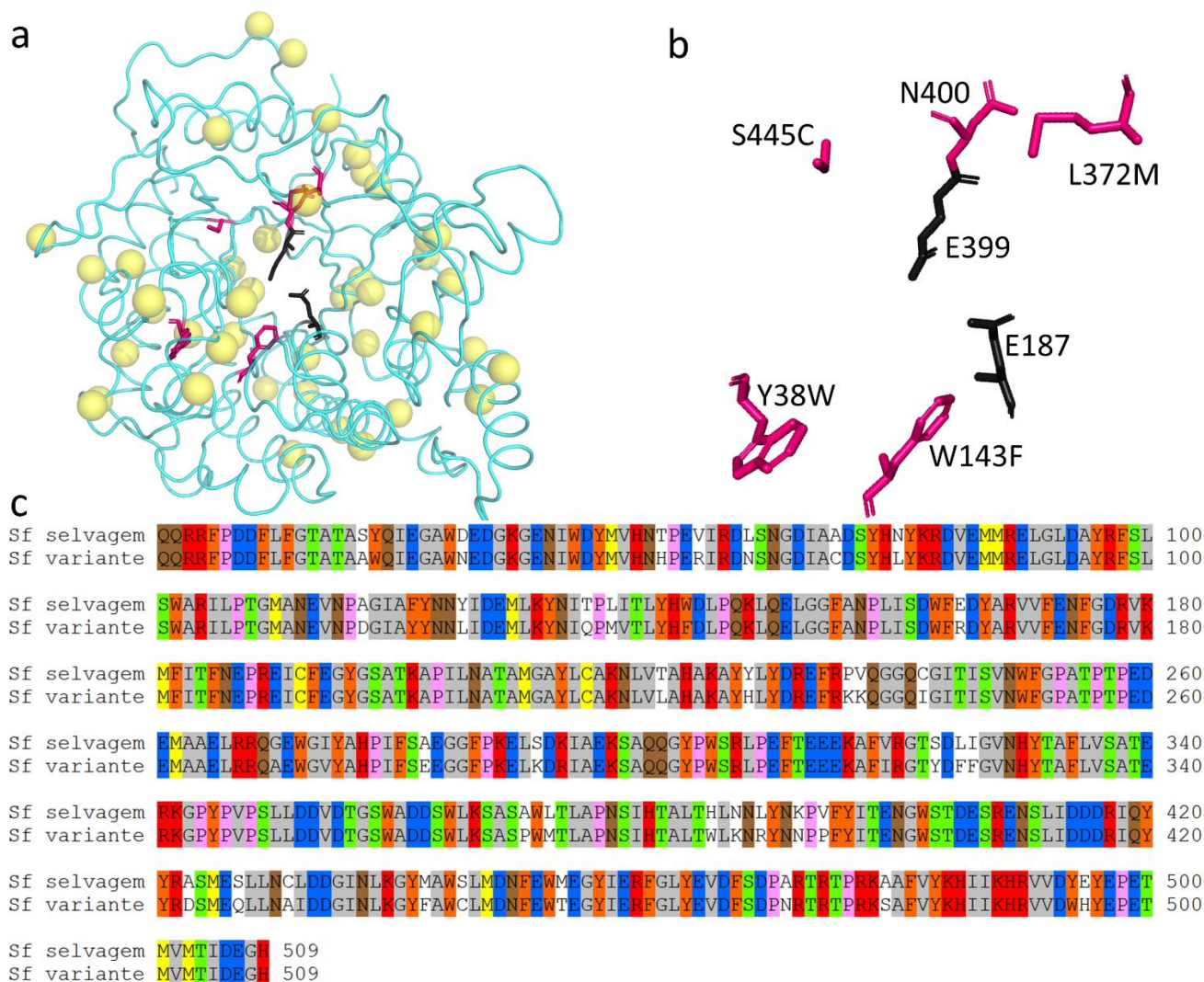


Figura 1.24. Variante de Sfβgly gerada pelo PROSS e FuncLib. **a:** Modelo da Sfβgly variante, destacando as posições em esferas amarelas que foram modificadas pelo PROSS para aumentar a estabilidade da enzima. **b:** Destaque em roxo dos resíduos ao redor do sítio ativo que foram mutados usando o FuncLib para ampliar o repertório de substrato, e em preto são destacados os resíduos catalíticos que não foram modificados. **c:** Alinhamento de sequência da Sfβgly selvagem e variante.

O próximo passo foi solicitar a construção do gene que codifica a variante Sfβgly_PROSS_FuncLib em plasmídeo de expressão, produzi-la em bactérias e purificá-la para verificar suas características enzimáticas e oligoméricas. O seu perfil cromatográfico em SEC indicou que a variante é um monômero em solução (**Figura 1.25a**), uma surpresa, já que a interface do homodímero não foi alvo de mutações nem do PROSS e do FuncLib. É provável que as mutações introduzidas

nesta variante tenham alterado profundamente a distribuição das populações conformacionais comparada à enzima selvagem, ao ponto de ter extinguido aquela conformação que possui mais afinidade para se dimerizar.

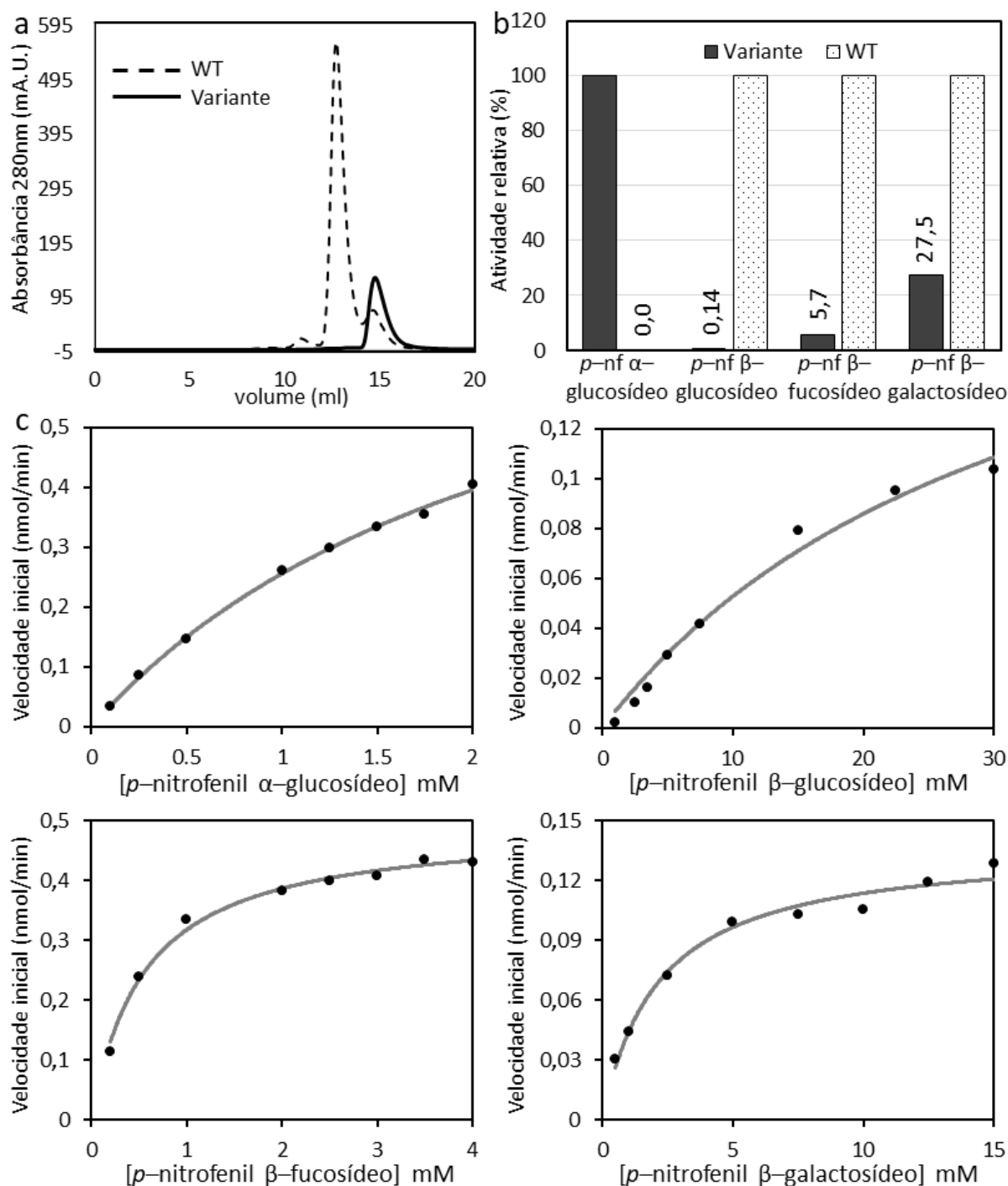


Figura 1.25. Experimentos de caracterização da variante Sfbgly_PROSS_FuncLib purificada. **a:** Perfil de SEC da variante (linha contínua) e da selvagem (WT, linha pontilhada). **b:** Comparação da atividade relativa entre substratos sintéticos da variante com a selvagem. **c:** Determinação dos parâmetros de cinética enzimática da variante com diferentes substratos sintéticos. Linha contínua é o ajuste da equação de Michaelis–Mentem sobre os dados.

Para nossa surpresa, os testes de ensaio enzimático revelaram a capacidade da variante de hidrolisar tanto o substrato *p*-nitrofenil β- e α-glucosídeo, uma habilidade rara entre glicosidases, que geralmente se especializam por um tipo de anômero. Na **Figura 1.25b,c** e **Tabela 1.4** são mostradas as comparações dos parâmetros cinéticos obtidos para a variante Sfbgly_PROSS_FuncLib e a enzima selvagem (WT) usando substratos sintéticos. Nota-se que o ganho de função de hidrólise do α-glucosídeo não aboliu a atividade pelos outros β-substratos, porém uma diminuição de eficiência catalítica generalizada é evidente. Na variante, a maior especificidade, expressa no k_{cat}/K_M , continua sendo pelo substrato *p*-nitrofenil β-fucosídeo (**Tabela 1.4**), assim como na selvagem. Porém, a especificidade pelo substrato *p*-nitrofenil β-glucosídeo reduziu-se em cerca de 730 vezes comparado ao k_{cat}/K_M da selvagem, até o ponto que a especificidade pelo substrato *p*-nitrofenil β-galactosídeo na variante se tornou maior do que pelo *p*-nitrofenil β-glucosídeo. Um claro contraste em relação à enzima selvagem.

Finalmente, o k_{cat}/K_M da enzima variante para o substrato *p*-nitrofenil α-glucosídeo é cerca de 40% daquele para o seu substrato de maior especificidade, *p*-nitrofenil β-fucosídeo, o que ressalta a alteração de estereo-especificidade promovida pelas múltiplas mutações introduzidas na variante Sfbgly_PROSS_FuncLib.

Tabela 1.4. Parâmetros de cinética enzimática da variante Sfbgly_PROSS_FuncLib purificada com substratos sintéticos comparados com a enzima selvagem (WT). Dados são médias ± SEM. “*p*-nf” abrevia *p*-nitrofenil.

	Variante		WT		Variante		WT	
	K_M (mM)		k_{cat} (s ⁻¹)		k_{cat}/K_M (mM ⁻¹ s ⁻¹)			
<i>p</i> -nf α-glucosídeo	2,4 ± 0,2	–	0,15 ± 0,01	–	0,06 ± 0,01	–		
<i>p</i> -nf β-glucosídeo	33,6 ± 10,4	1,8 ± 0,2	0,04 ± 0,03	1,4 ± 0,1	0,0011 ± 0,0006	0,8 ± 0,1		
<i>p</i> -nf β-fucosídeo	0,6 ± 0,1	1,0 ± 0,1	0,08 ± 0,01	2,8 ± 0,1	0,15 ± 0,02	2,6 ± 0,3		
<i>p</i> -nf β-galactosídeo	2,1 ± 0,3	8,1 ± 0,9	0,023 ± 0,001	0,33 ± 0,03	0,011 ± 0,002	0,04 ± 0,01		

Para entender os fatores moleculares que permitem à variante Sfbgly_PROSS_FuncLib o uso de α-substrato, o AlphaFold2 foi mais uma vez utilizado para prever sua estrutura tridimensional e compará-la com a Sfbgly

selvagem. Na **Figura 1.26a** é mostrada a sobreposição da variante com a selvagem com um RMSD entre todos os pares de átomos igual a 0,85 Å. De fato, nota-se pouca diferença na estrutura do *backbone* em geral, a não ser por alguns *loops* que não se sobrepõem precisamente, como por exemplo no *loop* compreendido entre A338 até P348 (RMSD = 3,44 Å).

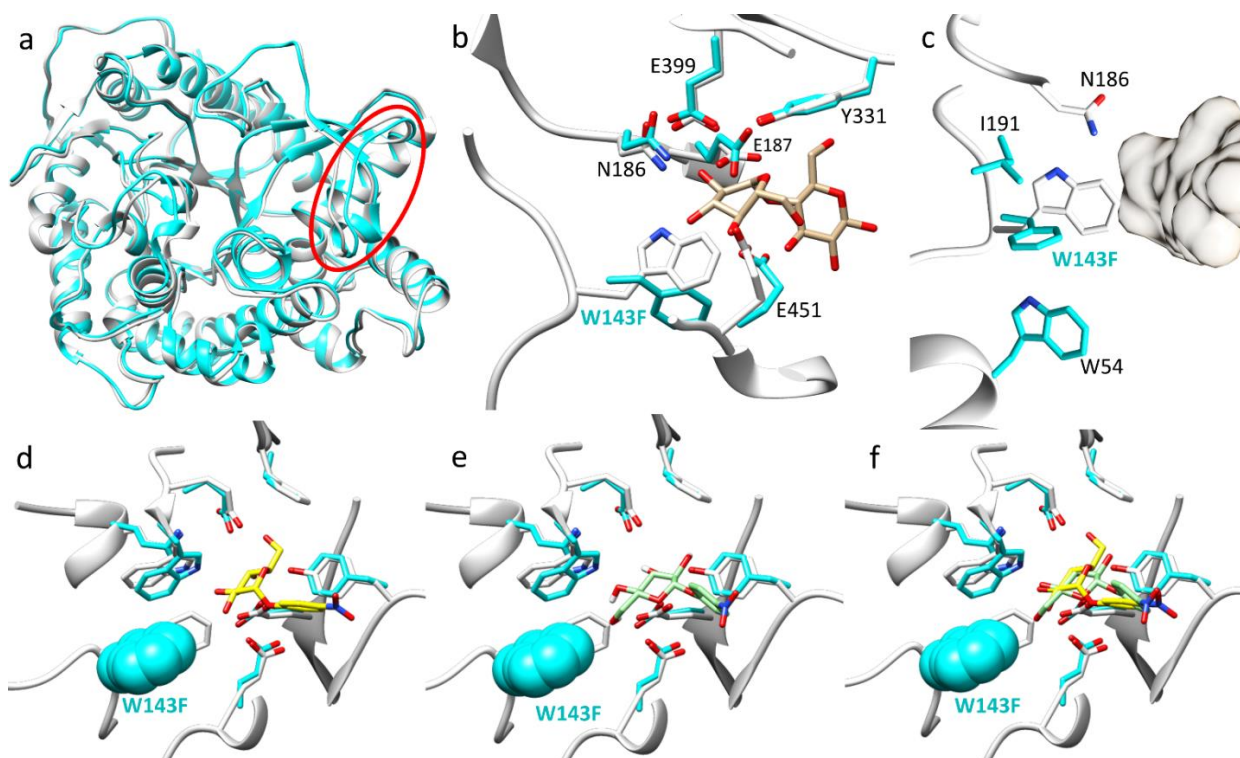


Figura 1.26. Comparação estrutural entre variante Sfbgly_PROSS_FuncLib (em ciano) e Sfbgly selvagem (em branco). **a:** sobreposição do modelo gerado por AlphaFold2 da variante com a estrutura cristalográfica da selvagem; a elipse vermelha indica um local de pouca sobreposição. **b:** Destaque dos resíduos no sítio ativo com a presença do substrato celobiose em bege. **c:** Visão do sítio ativo com a celobiose representada em superfície molecular. **d, e, f:** Sítio ativo da variante com *p*-nitrofenil α -glucosídeo (amarelo), *p*-nitrofenil β -glucosídeo (verde) e a sobreposição dos dois substratos, respectivamente. O *docking* automático dos substratos foi feito com AutoDock Vina (Trott & Olson, 2009) do Chimera.

A diferença estrutural que pode ser crucial para a nova estereoespecificidade da variante é a mutação W143F (**Figura 1.26b,c**). Com esta mutação, fica claro que há mais espaço no sítio ativo, facilitando a ligação e posicionamento do α -glucosídeo. Nas **Figura 1.26d–f** também é mostrado por *docking* que os substratos se ajustam bem no sítio ativo na enzima variante, e que ambos os substratos ocupam a mesma região. Além disso, os resíduos catalíticos mantêm curtas distâncias em relação ao carbono 1 e o oxigênio da ligação glicosídica

(Figura 1.27), de acordo com uma referência observada da estrutura cristalográfica de uma β -glicosidase GH1 de *Neotermes koshunensis* com o substrato *p*-nitrofenil β -glucosídeo (Jeng et al., 2011).

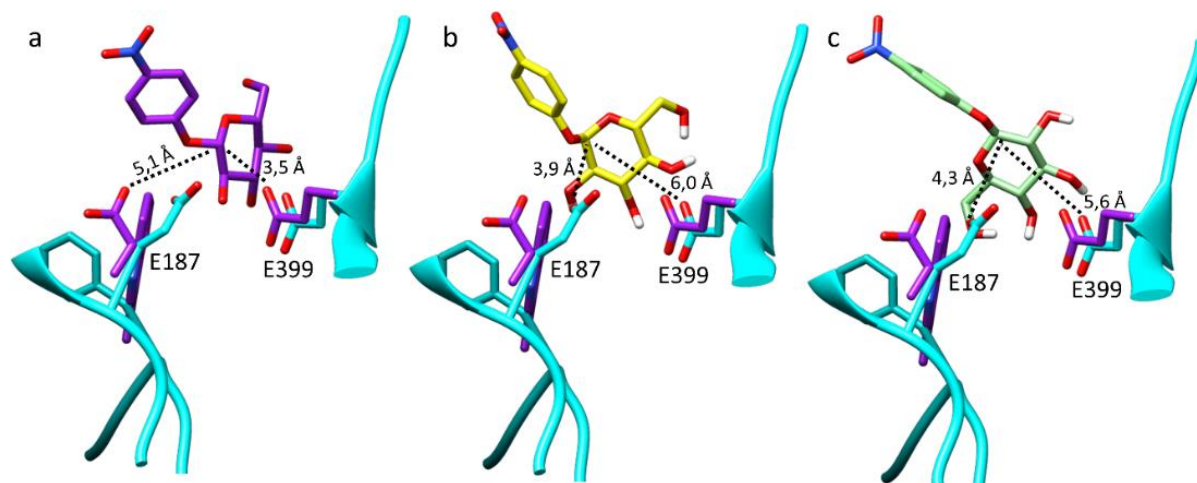


Figura 1.27. Distâncias interatômicas entre resíduos catalíticos e o carbono 1 do substrato em β -glicosidases. Para referência, a estrutura cristalográfica da β -glicosidase GH1 de *Neotermes koshunensis* (pdb 3AI0; em violeta) que contém o substrato *p*-nitrofenil β -glucosídeo foi sobreposta à estrutura da variante Sfbgly_PROSS_FuncLib (ciano). **a:** Distâncias entre átomos na estrutura da referência. **b, c:** distâncias entre átomos dos resíduos da variante com os substratos α -glucosídeo (amarelo), β -glucosídeo (verde) posicionados por Autodock Vina (Trott & Olson, 2009).

De qualquer modo, apesar da confiabilidade da predição da estrutura e a análise do *docking* feita com a variante de Sfbgly, é precoce isolar as mutações que foram capazes de alterar a estereo-especificidade da enzima. É possível, inclusive, que nesta variante, com o acúmulo de mais de 40 mutações, a nova função seja resultado da epistase (Noor et al., 2012). Realmente, o desvio da função de uma enzima que foi refinada ao longo de anos pela evolução é contrabalanceado pela tendência de manter a estabilidade do enovelamento da proteína (Tokuriki et al., 2008). Para aliviar os efeitos desestabilizadores de mutações cruciais para ganho de função, estratégias como as usadas na construção da variante Sfbgly_PROSS_FuncLib podem ser um ponto de partida, além de outros meios já validados, como por exemplo, o uso de proteínas com enovelamento de alta estabilidade (Teufl et al., 2022). Além disso, em termos de eficiência catalítica, a nova função da variante (Tabela 1.4) é cerca de 11 vezes menor comparada à eficiência da enzima selvagem com o substrato natural (Tabela 1.1), o que nos leva a questionar se a nova atividade forneceria suficiente vantagem adaptativa ou em

algum processo biotecnológico. Vale notar que em situações em que se deseja aumentar a eficiência catalítica são indicadas estratégias de evolução dirigida (Trudeau & Tawfik, 2019).

1.4 – Conclusões e perspectivas

Os experimentos realizados com a β-glicosidase Sfβgly da *Spodoptera frugiperda* revelaram duas características marcantes: (i) existe uma conexão entre o seu estado oligomérico e sua atividade enzimática, (ii) a versatilidade da estereoespecificidade pelo substrato resultante da inserção de múltiplas mutações na enzima selvagem.

A Sfβgly selvagem pode estar no estado de dímero ou monômero, este último com menor eficiência catalítica comparado ao dímero. Esta característica, uma enzima com múltiplas formas oligoméricas com diversidade funcional, é explicada pela dinâmica de sua estrutura tridimensional, resultando em populações de conformações distintas. Além disso, a cinética de dissociação do dímero sugere que no caso de Sfβgly o mecanismo de ligação entre subunidades é por via de seleção de confôrmeros.

Testes com mutações na interface do dímero de Sfβgly sugerem que este estado é marginalmente estável, já que mutações simples são capazes de impedir a formação do homodímero. No entanto, mutantes na interface planejadas para estabilizar o dímero causaram insolubilidade da enzima, sugerindo que a interface é um local com potencial para causar agregação, principalmente se há um incremento de resíduos hidrofóbicos.

A Sfβgly tem potencial para diversificar seu repertório de substratos através de *design* por métodos computacionais, como foi visto na variante Sfβgly_PROSS_FuncLib, que é capaz de utilizar um α- substrato. A variante também demonstrou que a forma dimérica pode ser desestabilizada sem alterar resíduos na interface, possivelmente por remodelar a população de conformações da enzima e tornar inacessível a que possui maior afinidade entre monômeros. Uma opção não testada nesta tese para estabilizar o dímero de Sfβgly, também válida para outras proteínas, seria guiar a formação de estrutura quaternária

através da estabilização de conformações de maior afinidade entre monômeros ao invés de simplesmente aumentar hidrofobicidade na interface.

Métodos computacionais automáticos como os utilizados para o *design* da variante Sfbgly_PROSS_FuncLib são viáveis para promover ganho de função em enzimas. Passos discretos de ganho de função, aumentando a estabilidade da proteína, parecem ser uma forma lógica e segura para produzir enzimas que catalisem reações ainda não existentes na natureza. Porém, o preciso refino da nova função ainda é um fator limitante que não é trivialmente realizado por métodos computacionais. Uma maneira robusta de implementar estabilidade e novas funções é através de inferência de sequências ancestrais (Trudeau & Tawfik, 2019), o que será explorado e discutido no próximo capítulo.

1.5 – Publicação em jornal científico

Received: 12 May 2020 | Revised: 22 June 2020 | Accepted: 23 June 2020
DOI: 10.1002/pro.3908



ARTICLE



Homodimerization of a glycoside hydrolase family GH1 β -glucosidase suggests distinct activity of enzyme different states

Felipe A. M. Otsuka | Rafael S. Chagas | Vitor M. Almeida | Sandro R. Marana

Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, São Paulo, Brazil

Correspondence

Sandro R. Marana. Av. Prof. Lineu Prestes, 748 – São Paulo, SP 05508-000, Brazil.
Email: smarana@iq.usp.br

Funding information

Conselho Nacional de Desenvolvimento Científico e Tecnológico; Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Grant/Award Number: Finance Code 01; Fundação de Amparo à Pesquisa do Estado de São Paulo, Grant/Award Numbers: 2016/22365-9, 2018/18537-4, 2018/25952-8

Abstract

In this work, we investigated how activity and oligomeric state are related in a purified GH1 β -glucosidase from *Spodoptera frugiperda* (Sfβgly). Gel filtration chromatography coupled to a multiple angle light scattering detector allowed separation of the homodimer and monomer states and determination of the dimer dissociation constant (K_D), which was in the micromolar range. Enzyme kinetic parameters showed that the dimer is on average 2.5-fold more active. Later, we evaluated the kinetics of homodimerization, scanning the changes in the Sfβgly intrinsic fluorescence over time when the dimer dissociates into the monomer after a large dilution. We described how the rate constant of monomerization (k_{off}) is affected by temperature, revealing the enthalpic and entropic contributions to the process. We also evaluated how the rate constant (k_{obs}) by which equilibrium is reached after dimer dilution behaves when varying the initial Sfβgly concentration. These data indicated that Sfβgly dimerizes through the conformational selection mechanism, in which the monomer undergoes a conformational exchange and then binds to a similar monomer, forming a more active homodimer. Finally, we noted that conformational selection reports and experiments usually rely on a ligand whose concentration is in excess, but for homodimerization, this approach does not hold. Hence, since our approach overcomes this limitation, this study not only is a new contribution to the comprehension of GH1 β -glucosidases, but it can also help to elucidate protein interaction pathways.

KEYWORDS

conformational selection, enzyme activity, glycoside hydrolase, homodimer, oligomerization, *Spodoptera frugiperda*, β -glucosidase

1 | INTRODUCTION

Glycoside Hydrolases are enzymes that are expressed among life forms from bacteria to animals and are involved in a series of biological demands, such as digestion, availability of cellular free glucose and post-translation modification.^{1–3} Glycoside hydrolase (GH) cataloged in the CAZY databank

(www.cazy.org), displays more than 150 families based on sequence and structural similarities. The GH1 family comprises β -glycosidases, enzymes that hydrolyze the O- or S-glycosidic linkages of β -glycosides, and folds as a TIM-barrel ($[\beta/\alpha]_8$ barrel),⁴ a repetition of a (β/α) motif.⁵

For many years, GH1 β -glucosidases have been studied regarding their physiological function, substrate specificity,

Capítulo 2

Mutações de resistência e evolução do domínio catalítico tirosina quinase Abl1 humana (Abl1-KD)

2.1 – Introdução

2.1.1 – Resistência no tratamento de câncer: o caso da leucemia mieloide crônica e o oncogene BCR–Abl1

O câncer ameaça a vida de dezenas de milhões de pessoas no mundo (Bray et al., 2013) e mesmo com os grandes avanços no seu tratamento e prevenção descobertos nas últimas décadas, esforços para domá-lo continuam sendo crescentes e passando cada vez mais a ter ênfase em iniciativas multidisciplinares (Umar et al., 2012). O entendimento do câncer como doença neoplásica pode apoiar-se em pelo menos seis características distintas (**Figura 2.1**) (Hanahan & Weinberg, 2011) que levam as células tumorais a romper com a homeostase do organismo (Bruce Alberts et al., 2009).

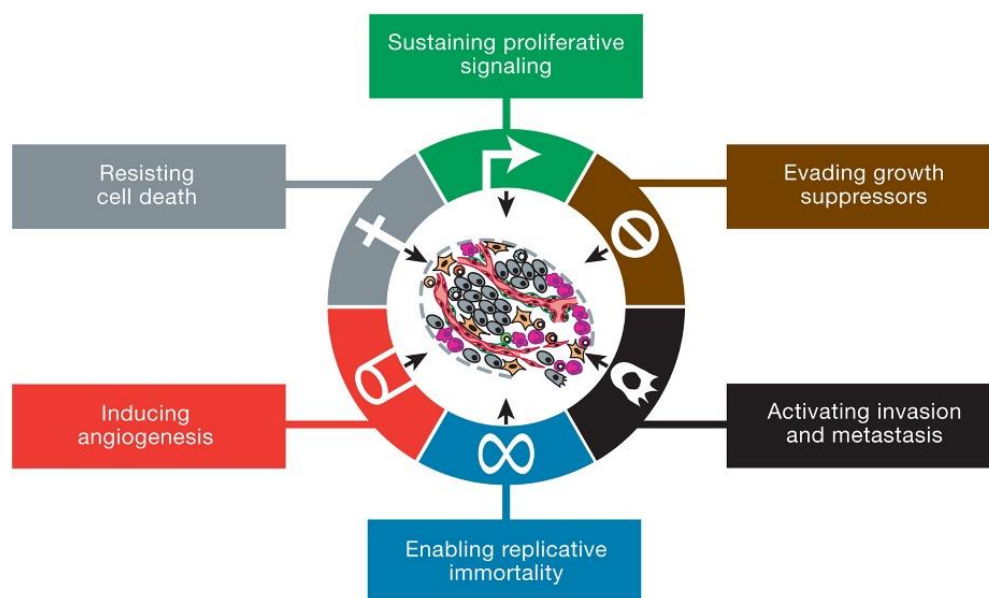


Figura 2.1. Características que permitem ao tumor cancerígeno crescer e sofrer metástase disseminativa. Figura retirada de Hanahan D. e Weinberg R. (Hanahan & Weinberg, 2011).

Dentre estas características, o desenvolvimento do câncer pode resultar particularmente de células nas quais a capacidade de sustentar a proliferação

crônica deriva da desregulação de proteínas envolvidas no ciclo celular, como por exemplo as enzimas tirosinas quinases (Hanahan & Weinberg, 2011). Isto é o que acontece na Leucemia Mieloide Crônica (CML), causada por uma transformação neoplásica em células-tronco hematopoiéticas (Deininger et al., 2005), na qual o gene *breakpoint cluster region* (BCR) do cromossomo 22 é translocado em t(9;22)(q34;q11), justaposto ao gene Abelson Leukemia (Abl) no cromossomo 9 (Ren, 2005). A fusão BCR-Abl1 resulta em uma enzima constantemente ativa, por isso o domínio tirosina quinase Abl1-KD (UniProt ID: P00519) é um importante alvo terapêutico e usada no *design* de moléculas inibidoras. Dentre estas, são mostradas na **Figura 2.2** as drogas de primeira escolha aprovadas para o tratamento de CML: Imatinib, Nilotinib, Dasatinib, Bosutinib e Ponatinib que atuam competindo com o ATP pela ligação no Abl1-KD (Hochhaus et al., 2020). Inicialmente, o tratamento com inibidores de Abl1-KD se mostrou muito eficiente no combate da CML, o que resultou em um grande entusiasmo. No entanto, foi observado que alguns pacientes sofrem com o relapso da doença, que é acompanhado por resistência ao tratamento com inibidores de tirosina quinase (Braun et al., 2020), o que realça a visão do câncer como um sistema complexo e adaptativo.

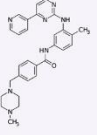
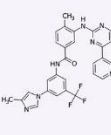
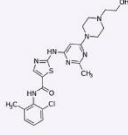
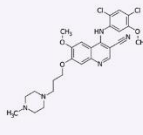
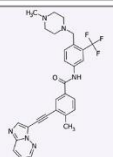
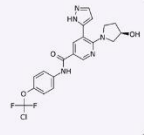
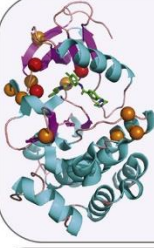
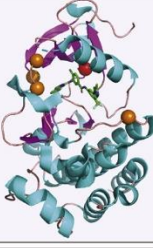
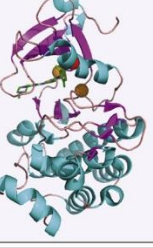
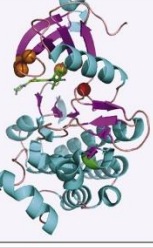
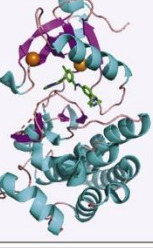

Inhibitor	Imatinib	Nilotinib	Dasatinib	Bosutinib	Ponatinib	Asciminib
Chemical Structure						
Crystal Structure						
Binding Conformation	Inactive	Inactive	Active	Both	Inactive	Myristoyl Pocket
Resistance	Y253 Q252 E255 F317 T315 M351 M244 M355 L248 F359 G250 H396	T315 L248 Y253 E255 F359	T315 V299 F317	T315 V299 L248 G250 E255 F317	T315 E255	A337 W464 P465 V468 I502

Figura 2.2. Inibidores de tirosina quinase da Abl1-KD e suas respectivas estruturas químicas ligadas na estrutura tridimensional do domínio catalítico. Figura originalmente publicada em Braun, T.P., et al. 2020.

A dinâmica do desenvolvimento e resistência do câncer ao tratamento com inibidores de tirosina quinase é bem compreendida pela teoria da evolução clonal (*clonal evolution*), que descreve o câncer como um processo inerentemente evolutivo, de circunstâncias únicas para cada clone que acumula mutações de forma sequencial (**Figura 2.3**) (Greaves & Maley, 2012; Nowell, 1976). O tratamento da CML com o uso de inibidores da tirosina quinase Abl1-KD, por exemplo, causa a morte de uma população celular com fenótipo BCR-Abl1+, significando uma forte pressão seletiva favorável a células com genes de resistência (Braun et al., 2020; Schmitt et al., 2018).

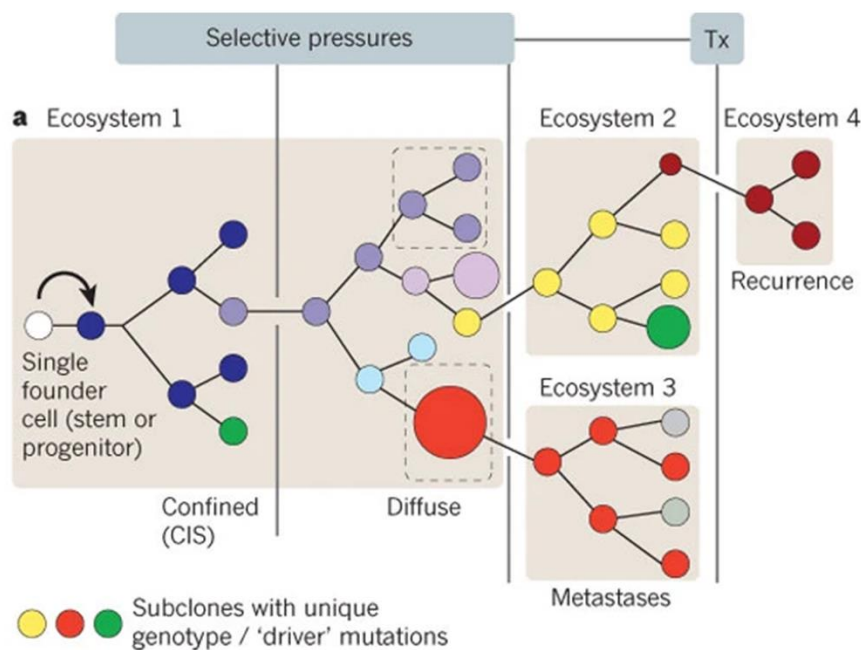


Figura 2.3. Evolução clonal de células cancerígenas em diferentes contextos de pressão seletiva e terapias (Tx) ou microambientes (*Ecosystem*). Cada círculo colorido representa um clone geneticamente distinto. Um paralelo da resistência e relapso da CML em pacientes é representado por clones no *Ecosystem 2* e *4*. Figura originalmente publicada em Greaves M. e Maley C.C. 2012 (Greaves & Maley, 2012).

2.1.2 – Mutações de resistência e conformação da Abl1-KD

Cerca de 25% dos pacientes com CML desenvolve algum tipo de resistência aos inibidores de quinase causada por mutações no domínio Abl1-KD (O'hare et al., 2012). Estruturalmente, este domínio possui uma dinâmica singular que gera

um conjunto de conformações biologicamente relevantes. Três delas se destacam: são as conformações inativa, ativa e uma intermediária entre as duas anteriores, chamada Src-like (nome dado por semelhança a quinase Src, homóloga ao Abl1). Na **Figura 2.4** são mostradas as estruturas de cada conformação, assim como a posição na sequência que estão associadas a conferir resistência aos inibidores após certas mutações. Através do R.M.S.D. dos carbonos alfas (Ca) dos resíduos representados em esferas amarelas, nota-se que relevantes mudanças conformacionais ocorrem na enzima, com ênfase em duas porções estruturais, uma α -hélice na extremidade C-terminal (**Figura 2.4a**) e o *loop* de ativação, mostrado em cinza, que contém o resíduo H396.

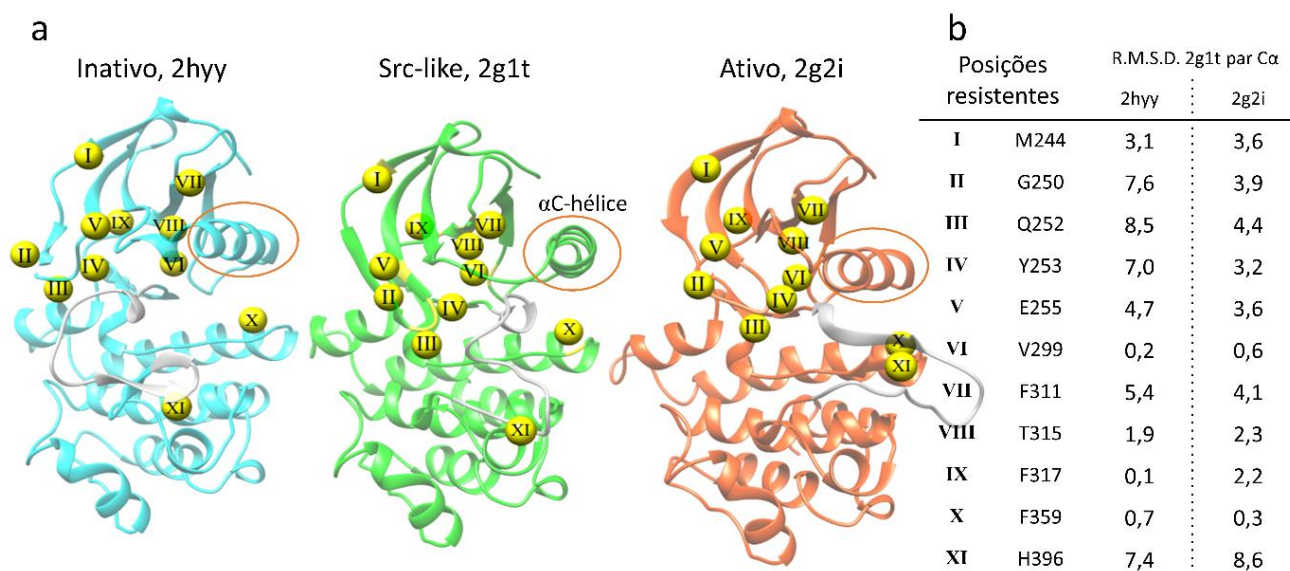


Figura 2.4. Conformações do domínio Abl1-KD (**a**; PDB ids.: 2hyy, 2g1t, 2g2i) e (**b**) posições que com certas mutações conferem resistência a inibidores de quinase. Os carbonos-alfas das posições estão destacadas em amarelo e com numeração romana. Duas estruturas importantes para a troca de conformação também são destacados: a hélice- α que é indicada por uma elipse vermelha e o *loop* de ativação que contém o domínio DFG conservado destacado na cor cinza. Figura originalmente de Otsuka F.A.M. e Bjelic S. (Otsuka & Bjelic, 2022).

A conformação inativa da Abl1-KD ganha atenção especial por ser o principal modelo estrutural para o desenho de inibidores, como foram nos casos do Imatinib, Nilotinib e Ponatinib. Estes são primariamente planejados para competir ao sítio ativo do ATP na Abl1-KD (**Figura 2.2**), que mesmo na forma inativa, tem afinidade para ligar ao ATP. O inicial sucesso no tratamento de CML

através destes medicamentos que ligam na forma inativa foi graças a capacidade de “travar” a conformação da Abl1-KD na forma inativa, impedindo a troca conformacional para a forma ativa. A conformação ativa desta enzima possui duas características estruturais bem conservadas chamadas de coluna regulatória (R-spine) e a coluna catalítica (C-spine) que quando organizadas da maneira correta, deixam a quinase “pronta para ação” (de Oliveira et al., 2016; Taylor et al., 2012).

Apenas um conjunto limitado de mutações é capaz de gerar o fenótipo de resistência na CML, o qual tem seu efeito por duas vias: dificultam a ligação do inibidor ao Abl1-KD através de impedimento estérico ou tornam mais estável a conformação ativa da enzima através de um remodelamento na superfície de energia desta conformação ou por alteração da sua dinâmica molecular (Lovera et al., 2015; Shah & Sawyers, 2003; Wilson et al., 2015). As mutantes pontuais que estão associadas ao relapso da doença em pacientes com CML e tem suas posições destacadas na **Figura 2.4** são: M244V, G250E, Q252H, Y253H, E255(K,M,V), V299L, F311I, T315(I,M), F317L, F359(V,C), H396(P,R) (Iqbal et al., 2013). Dentre essas substituições, a que se destaca é o T315I e duplos mutantes dela com outras mencionadas acima, que conferem resistência ao potente inibidor quinase ponatinib (Zabriskie et al., 2014). Isto destaca o problema que a resistência do câncer representa aos pacientes, pois, mesmo novos medicamentos desenhados e altamente eficientes estão susceptíveis a serem “enganados” e terem suas eficiências diminuídas por novos ciclos de pressão seletiva que culminam na proliferação de clones resistentes.

2.1.3 – Objetivos e hipóteses

A aquisição de resistência aos medicamentos não é um evento exclusivo do tratamento do câncer, também ocorrendo, por exemplo, no tratamento de infecções bacterianas (Merrih & Kohli, 2020; Vasan et al., 2019). Ambas as situações se desenrolam sob as regras da evolução, que é a força motriz que garante o surgimento da tolerância aos tratamentos medicamentosos. Portanto, conhecer e anteceder as consequências de pressões evolutivas feitas por medicamentos podem gerar vias para aprimorar os tratamentos já existentes. Desta forma, duas hipóteses foram formuladas. A primeira buscou verificar se algumas das mutações que geram

resistência em pacientes com CML causam algum impacto na estabilidade das conformações que Abl1-KD possui, com ênfase naquelas apresentadas na **Figura 2.4**. A segunda buscou avaliar se a seleção desses resíduos de resistência está submetida a um controle evolutivo predeterminado com base na sua ancestralidade. As hipóteses são:

- I- As possibilidades da superfície de energia em função de mutações (*mutational energy landscape*) para cada conformação de Abl1-KD (inativa, ativa e Src-like) destacam as substituições que conferem maior estabilidade para uma determinada conformação. Então mutações que conferem resistência devem em certo grau estabilizar a conformação ativa em relação às outras.
- II- As mutações de resistência são acessíveis dentro de uma faixa tolerável de energia conformacional e pressões evolutivas, ou seja, apenas uma parcela das 19 possibilidades de resíduos é acessível para troca em uma mesma posição. Contrário a este cenário, se as posições na sequência de Abl1-KD que conferem resistência acessam resíduos na mesma taxa da deriva genética neutra (*neutral genetic drift*), então não há restrições evolutivas para a possibilidade de resíduos pela Abl1-KD.

2.2 – Materiais e métodos

2.2.1 – Modelos estruturais e conformações da Abl1-KD

As estruturas cristalográficas de Abl1-KD com PDB IDs 2hyy, 2g1t e 2g2i foram editadas de modo a terem o resíduo W235 na sua extremidade N-terminal e o resíduo Q498 no C-terminal. As estruturas 2hyy e 2g2i tiveram os átomos ausentes modelados pelo Rosetta. O ajuste estrutural global foi feito com o Chimera (Pettersen et al., 2004) usando o algoritmo Needleman-Wunsch com a matriz BLOSUM-62 para calcular o R.M.S.D. entre os $C\alpha$.

2.2.2 – Refinamento estrutural de Abl1-KD nativo e mutado

O aplicativo RosettaScripts (Fleishman et al., 2011), parte do *software* de modelagem de proteínas Rosetta (Leaver-Fay et al., 2011), foi usado para o projeto, empregando o arquivo *.resfile* para especificar um resíduo de aminoácido particular em uma determinada posição na sequência. PackRotamersMover e

FastRelax (Khatib et al., 2011) foram usados com 5 iterações. Todos os 20 aminoácidos foram amostrados em cada posição da estrutura proteica, executando 35 trajetórias cada. Assim, um total de 184.800 modelos de sequência mutada e nativa foram calculados por conformação. A comparação da energia total e energia residual (em REU, do inglês *Rosetta Unit Energy*) entre conformações foi feita a partir da energia média e desvio padrão calculados a partir dos 5 menores valores. Para análise das conformações com a sequência nativa, foram modeladas 9.240 modelos por conformação com o Rosetta. Para avaliar a convergência mínima de nossos cálculos, dois conjuntos independentes de 15 trajetórias foram determinados para a estrutura 2hyy usando o mesmo protocolo FastRelax (Khatib et al., 2011). Como acima, a energia total e a média de energia residual individual dos 5 menores valores foram correlacionadas para avaliar a convergência.

2.2.3 – Otimização de sequência de Abl1-KD

Para projetar Abl1-KD nas conformações inativa, Src-like e ativa, foi usado um protocolo similar ao FastRelax, mas em vez de alterar um aminoácido por posição, apenas às 11 posições resistentes foi permitido mudar entre os 20 aminoácidos canônicos durante a mesma trajetória de cálculo. No total, foram calculadas 750 trajetórias por conformação. As 10 sequências de menor energia de cada conformação foram alinhadas e sua identidade de sequência foi analisada. Os resíduos em posições resistentes foram anotados.

2.2.4 – Reconstrução ancestral de Abl1-KD e *design* de ancestrais

A sequência Abl1-KD (UniProtKB - P00519) foi usada como base para identificar tirosina quinases homólogas (TK) no banco de dados Uniprot (Bateman et al., 2021). Foram listadas proteínas da família SH3-SH2-TK restritas ao subgrupo de Abl1, Src, Tec, Csk e delimitados dois grupos externos formados pela proteína quinase A (PKA) (Barker & Dayhoff, 1982) e o receptor do fator de crescimento de fibroblastos, totalizando 41 sequências. As sequências foram alinhadas no servidor *web* GUIDANCE2 (Sela et al., 2015) com algoritmo MAFFT (Katoh & Standley, 2013) seguido de inferência de árvore filogenética usando o *software* MEGA-X versão 10.1.7 (Kumar et al., 2018) com o método de máxima

verossimilhança e parâmetros do modelo de substituição LG+G+I com 1.000 réplicas de *bootstraps*. As sequências ancestrais foram inferidas usando PhyML versão 3.3 (Guindon et al., 2010). As matrizes de pontuação específicas de posição (PSSMs), extraídas com um *script* escrito em *python* a partir da reconstrução ancestral, foram usadas para determinar a variabilidade de resíduos por posição.

Para inferir a variabilidade de resíduos durante a reconstrução ancestral, foi utilizado o PSSM de cada nó ancestral. Foi contado o número de resíduos com probabilidade maior que zero. Três grupos de posições foram feitos: i) um conjunto conservado de 11 posições funcionais em sequências ancestrais que não mudam ao longo do tempo (estas retêm apenas uma identidade de aminoácido, conseqüentemente contada como 1); ii) as posições resistentes, e iii) um conjunto sem posições funcionais ou resistentes incluídas. A média e o erro padrão da média de todas as variabilidades de resíduos contados por posição para um determinado grupo foram plotados para todos os nós ancestrais.

Estruturas específicas de conformação de sequências ancestrais foram construídas com Modeller (Eswar et al., 2008). Ambas as conformações inativas (PDB ID 2hyy) e ativas (PDB ID 2g2i) foram usadas como modelos. Os modelos também foram otimizados em sequência utilizando RosettaScripts com movimentadores FastRelax e SeqprofConsensus (Fleishman et al., 2011). O desenho da sequência foi permitido naquelas posições onde a probabilidade foi inferior a 100% e os resíduos de amostragem com probabilidade superior a 0% de acordo com os PSSMs inferidos. Assim, mesmo resíduos com pequena probabilidade de ocupar uma posição (menos de 5%) tiveram a chance de serem incorporados na sequência final se sua energia fosse favorável. No total, foram simuladas 500 trajetórias. Nesta fase, cada nó ancestral tinha sequências distintas que foram otimizadas tanto na conformação inativa quanto na ativa. As 10 sequências de menor energia total por conformação foram alinhadas e a sequência de consenso foi extraída como a sequência ancestral final calculada pela Rosetta para um determinado nó.

2.2.5 – Expressão, purificação e ensaio enzimático dos ancestrais de Abl1-KD

Selecionamos o Abl1-KD nativo (UniProtKB - P00519, resíduos 229 - 511) e dois nós ancestrais, anc6 e anc8, para expressão heteróloga em *E. coli* para validar nossa análise filogenética. Cada nó ancestral tinha duas sequências correspondentes, uma inferida pelo PhyML e outra otimizada pelo Rosetta. Cada sequência de interesse foi sintetizada (GenScript, EUA) e clonada em pET29b(+) em fase com um marcador de afinidade hexa-histidina no C-terminal. Abl1-KD e ancestrais foram co-transformados com o plasmídeo pCDFDuet-1b codificando a fosfatase YopH (Seeliger et al., 2005), para antagonizar a atividade de Abl1-KD e minimizar a citotoxicidade para as células *E. coli* BL21 (DE3) Tuner. A expressão da proteína foi induzida com IPTG 1 mM sob agitação por 16 h a 20°C em meio Luria-Bertani (LB) com canamicina e estreptomicina (50 µg/ml cada). As células foram colhidas por centrifugação (2.300 g, 20°C, 20 min), ressuspensas em tampão de lise (50 mM de HEPES, pH 7,4, NaCl 500 mM, imidazol 10 mM, glicerol 20%, Triton X-100 1%, 2-mercaptoetanol 5 mM) e lisadas por sonicação, seguida de centrifugação (30 min centrifugação a 11.770 g) para separar os detritos celulares. O sobrenadante foi aplicado à coluna de afinidade Co2+ (HisPur™ Cobalt Resin, Thermo Scientific™), lavado 3 vezes com tampão de ligação (HEPES 50 mM pH 7,4, NaCl 500 mM, imidazol 10 mM, glicerol 5%, 2-mercaptoetanol 5 mM) e eluído com tampão de eluição (tampão de ligação com imidazol 300 mM). Traços de impurezas de proteína e YopH fosfatase foram ainda removidos por cromatografia de troca iônica usando uma coluna de troca aniônica (Resource Q ou HiTrap Q HP) usando um gradiente linear misturando tampão A (Tris-HCl 20 mM pH 8,0, glicerol 5%, DTT 1 mM, NaCl 50 mM) e tampão B (tampão A + NaCl 1M). As atividades enzimáticas das quinases purificadas foram determinadas usando o ADP-Glo Kinase Assay (Promega Biotech AB), seguindo o manual do fabricante. As quinases e substratos foram diluídos em tampão de quinase (Tris 40 mM pH 7,5, MgCl₂ 20 mM, BSA 0,1 mg/ml, DTT 50 µM). O ensaio foi realizado em 20 µL

de tampão de quinase contendo 3 ng de enzima purificada, ATP 30 μM ultrapuro e Abltide 10 μM ou 75 μM (SignalChem). As reações em quatro réplicas foram feitas em placas de ensaio de poliestireno de 96 poços de meia área de fundo plano branco opaco (Corning) a 25°C por 1 h e a luminescência foi medida usando o leitor de placas TECAN SPARK 10 M com um tempo de integração de 1 s. As velocidades de reação foram obtidas por taxas de conversão (μmol de ADP min^{-1} determinado a partir de uma curva de calibração de ADP de conversão de ATP–ADP) dividida pela concentração de enzima medida com Qubit™ (Thermo, EUA).

2.3 – Resultados e discussão

2.3.1 – Energias das conformações e de mutações em Abl1-KD

As três estruturas cristalográficas da Abl1-KD que representam seus estados inativo, Src-like e ativo (**Figura 2.4**) tiveram as suas energias* conformacionais determinadas através de simulações de interação dos resíduos com refinamento de estrutura usando o software Rosetta (Leaver-Fay et al., 2011). A distribuição de energias totais obtidos de 9.240 ciclos para cada conformação é mostrada na **Figura 2.5a**, e realça que as conformações das duas formas não ativas (Src-like e inativa) são mais estáveis, resultado que concorda com a visão de uma quinase altamente regulada (Soverini et al., 2018). Apesar de que não era esperado que a conformação Src-like possuísse uma energia muito menor quando comparada com as outras duas formas, tratando-se de uma conformação transiente entre a forma inativa e ativa (Narayan et al., 2020; Oruganti & Friedman, 2021). Então é provável que o *score* da conformação Src-like seja um artefato dependente do campo de força utilizado no procedimento do Rosetta. Mas não obstante esta conformação mantém seu apelo como um bom modelo para o desenho de inibidores.

Avaliando os pares de resíduos entre conformações (**Figura 2.5b**), seus *scores* de energia são moderadamente correlacionados (coeficiente *r* de Pearson para Src-like vs inativo = 0,77; inativo vs ativo = 0,69; Src-like vs ativo = 0,66). Significado que em geral, a energia de cada resíduo é semelhante entre as conformações. Mas alguns resíduos claramente não se correlacionam, como no

* As energias conformacionais são uma aproximação da energia livre e quanto menor essa pontuação mais estável é comparada ao estado desenovelado. A energia de conformação é a soma das energias de resíduos. A energia do Rosetta é obtida através do somatório de termos definidos em Alford R.F. et. al. 2017, *e.g.*, energia de atração, energia de repulsão, energia de solvatação implícita, ...

exemplo da correlação entre inativo vs ativo, onde os resíduos em posição de resistência G250, Y253 E255 e F311 possuem energias distintas. Com essa análise é possível identificar resíduos que mais divergem em energia entre conformações distintas.

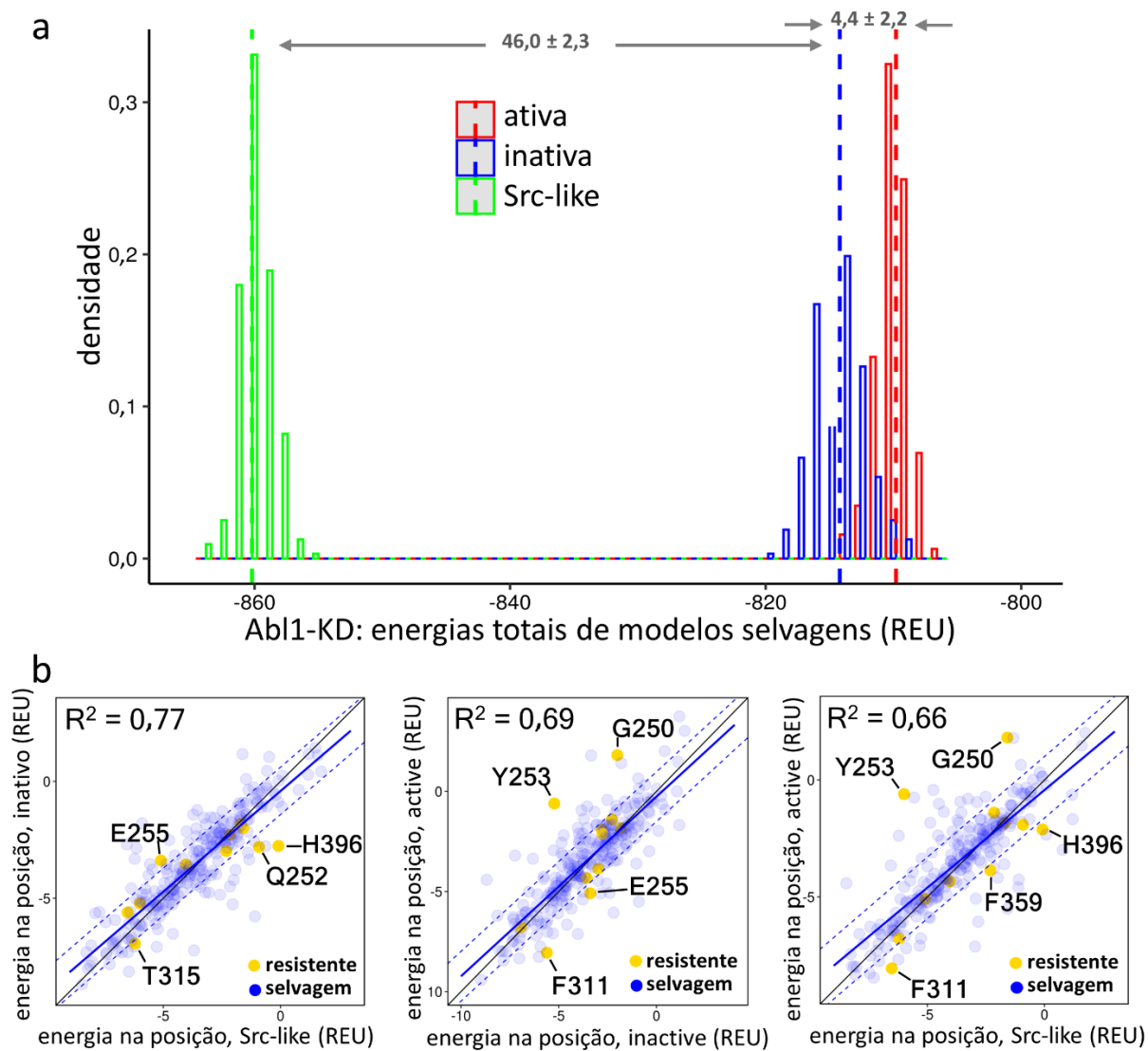


Figura 2.5. Amostragem de energia total das conformações de Abl1-KD com médias indicadas em linhas tracejadas (**a**) e correlação da energia de resíduo entre conformações (**b**); pontos amarelos indicam as posições de resíduos que conferem resistência quando substituídos, linha de tendência contínua em azul da correlação linear, e linha tracejada azul indica o erro padrão da estimativa. REU: Unidade de Energia do Rosetta.

Depois foram simulados e pontuados todos os possíveis mutantes simples para cada conformação da Abl1-KD. Suas energias totais se correlacionam fracamente entre as mutantes de cada conformação (**Figura 2.6a–c**) indicando que mutações pontuais são capazes de mudar a energia relativa entre conformações. Focando nas energias de resíduos em posições de resistência, as **Figuras 2.6d–e**

mostram as suas correlações entre conformações. Na comparação, por exemplo, entre a energia de resíduos nas conformações ativa e inativa (**Figura 2.6d**) nota-se que um *cluster* de mutações ditos resistentes, compostos por G250E, Q252H, E255M, E255V, F359C, e F359V, aparece com energias mais favoráveis para a forma ativa, o que concorda com um fenótipo de resistência causado pela maior estabilidade da conformação da enzima ativa. Porém, a mutação T315I, que é a principal causadora de resistência a inibidores, não tem sua energia de resíduo alterada significativamente, o que contrasta com a hipótese de estabilização da conformação ativa.

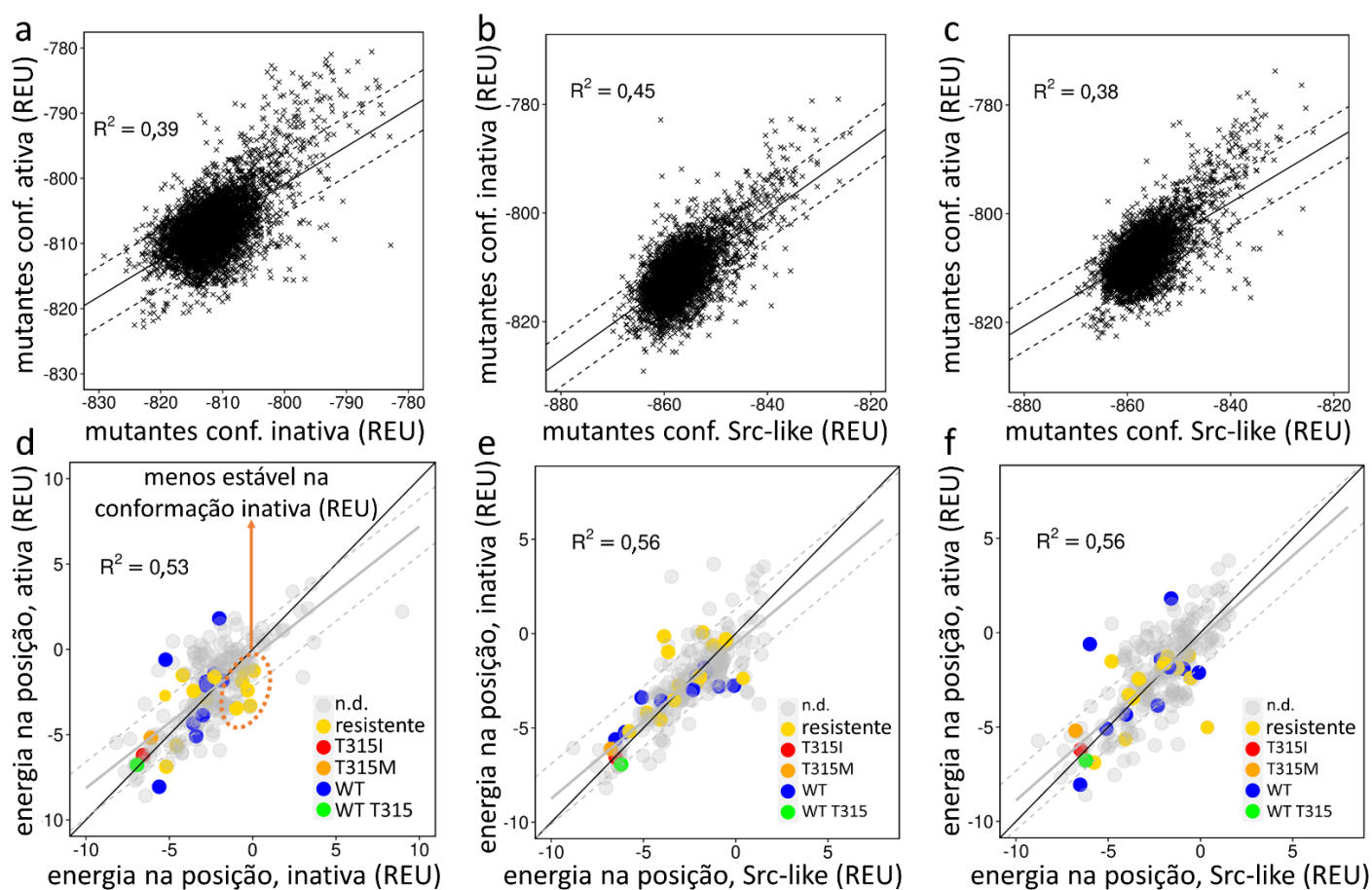


Figura 2.6. a–c: Correlação de energia total de todos os possíveis mutantes simples entre as conformações; linha de tendência contínua da correlação linear, e linha tracejada indica o erro padrão da estimativa. d–f: Correlação de energia do resíduo de todos os possíveis 20 resíduos em posições de resistência; linha de tendência contínua cinza da correlação linear, e linha tracejada cinza indica o erro padrão da estimativa; pontos em cinza indicam mutações não denominadas, em amarelo as de resistência, em vermelho e laranja as mutações na posição T315, azul e verde indicam resíduos selvagens.

Para buscar capturar a forma como as substituições em posição de resistência afetam as conformações em relação aos resíduos selvagens,

determinamos a diferença relativa de energia conformacional. Na **Figura 2.7a,c,e** são mostradas as energias totais de mutantes na forma inativa e ativa e as respectivas diferenças de energias entre elas. Nesta análise os mutantes resistentes M244V, G250E, Q252H, Y253H, F359VC e H396P remodelam a estabilidade da Abl1-KD, tornando a conformação ativa mais estável. No entanto, a **Figura 2.7e** também destaca que a média de energia das mutantes nas referidas posições quase não altera a relação de energia (barras vermelhas) comparada a da sequência selvagem (barra e linha pontilhada azul), com exceção das posições G250 e F359, que se substituídas tem altas chances de estabilizar a conformação ativa. Controversamente, as mutações resistentes nas posições E255, V299, F311, T315, F317 e H396 tornaram a conformação inativa mais estável (**Figura 2.7e**, sem setas vermelhas), sugerindo que a forma como estão relacionadas com a resistência aos inibidores é outro que a estabilização da forma ativa.

Nas **Figuras 2.7b,d,f** são destacadas as energias dos resíduos em posições de resistência nas conformações inativa e ativa, como também o quanto as substituições são alteradas energeticamente comparadas aos resíduos selvagens. Com este modo de análise, as substituições G250E, Q252H, Y253H, E255V, V299L, F359(C,V) e H396P são energeticamente mais favoráveis na conformação ativa (setas vermelhas, **Figura 2.7f**). E como visto anteriormente, as mutações de resistência na posição T315 não alteram sua energia quando comparadas com o resíduo selvagem. Além disso, mutações nessa posição se mostraram toleráveis para a conformação inativa. Desta forma ficou demonstrado que mutações nas posições de resistência são capazes de alterar a energia conformacional da Abl1-KD, o que pode ser atribuído primariamente pela instabilidade que essas substituições possuem na conformação inativa. No entanto, há algumas exceções, como por exemplo o mutante M244V, que tem a sua energia de resíduo praticamente inalterada comparando ao selvagem (**Figura 2.7f**), mas altera a relação de energia conformacional a favor da forma ativa (**Figura 2.7e**).

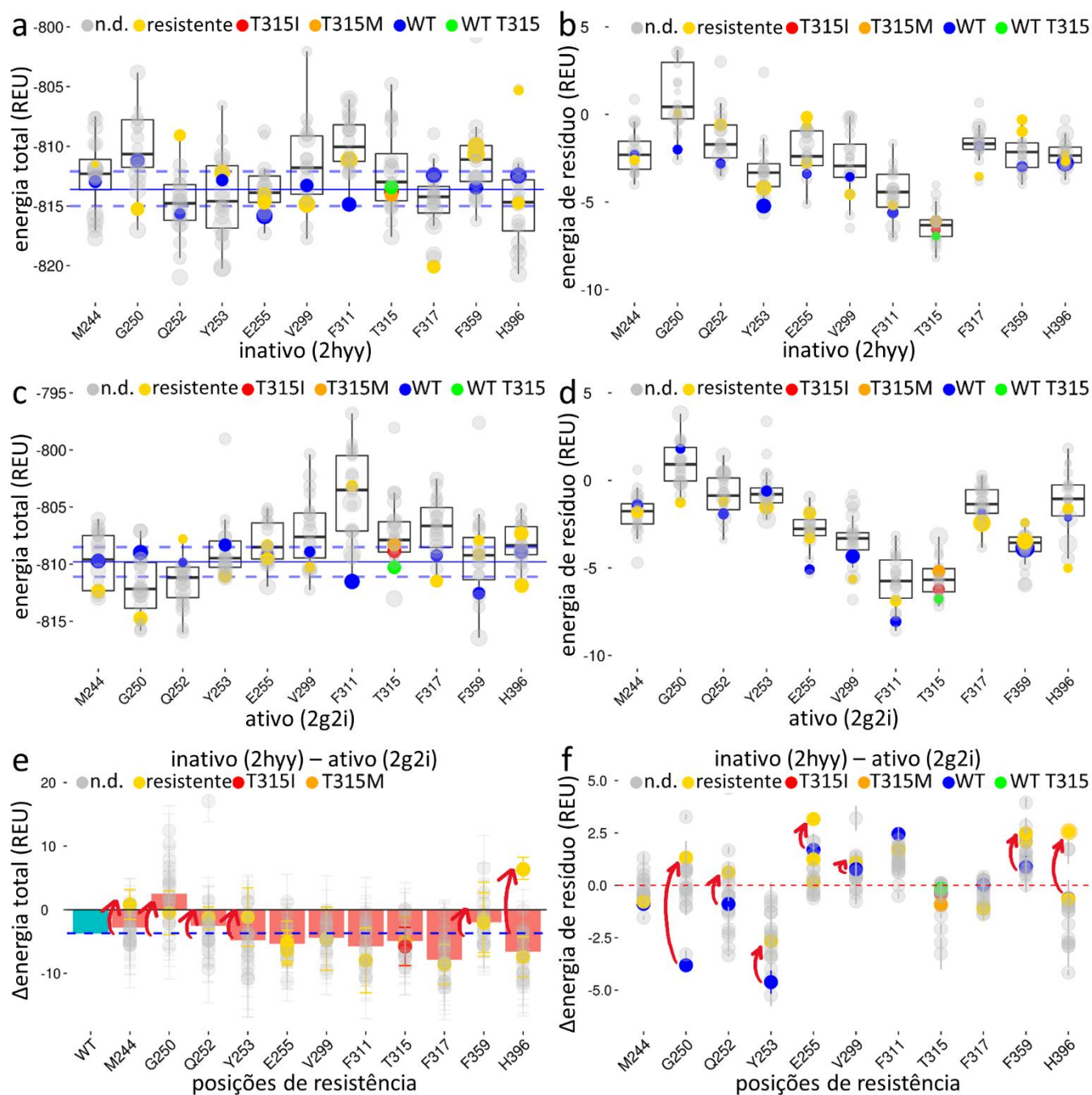


Figura 2.7. Análise de energia total (gráficos a esquerda) e de resíduo (gráficos a direita) após mutagênese de saturação em posições de resistência. **a–d:** Energia total e de resíduo das conformações inativa (**a, b**) e da ativa (**c, d**). **e:** Energia total relativa (inativo – ativo); setas vermelhas indicam mudanças na energia que estabilizam a conformação ativa após a mutação resistente; barras vermelhas representam a média da energia total relativa entre todos os possíveis mutantes na posição e a linha azul faz referência a diferença de energia (a mesma na **Figura 2.5a**) entre a conformação inativa e ativa de sequência selvagem (WT). **f:** Energia de resíduo relativa (inativo – ativo); setas vermelhas indicam mudanças na energia que estabilizam o resíduo na conformação ativa após a mutação resistente; pontos em cinza indicam mutações não denominadas. O tamanho dos círculos indica a magnitude do desvio padrões.

2.3.2 – Otimização de sequência dependente de conformação da Abl1-KD

Outra maneira de entender a correlação entre estabilidade da conformação e dos resíduos na estrutura é otimizando a sequência de aminoácidos em cada conformação. Neste tipo de protocolo, o Rosetta calcula os resíduos que melhor estabilizam a estrutura fornecida como *input* e devolve uma nova sequência. Com essa estratégia de otimização, buscou-se verificar qualitativamente o efeito da conformação na escolha de mutações em posições de resistência, ou seja, uma lógica inversa daquela realizada anteriormente, onde exaustivamente calculou-se o efeito de cada mutação sobre a estabilidade da conformação, e que ainda também dispensa grandes análises estatísticas.

As 12 posições de resistência foram selecionadas para que o Rosetta durante cada ciclo de otimização de estrutura, de um total de 750 por conformação, decidisse qual a combinação de aminoácidos (entre os 20 disponíveis) tornaria a estrutura conformacional mais estável. Em outras palavras, as sequências de menor energia seriam aquelas que minimizam a energia da conformação fornecida como *input*. Na **Tabela 2.1** são mostradas as mutações das sequências minimizadas para cada uma das três conformações que foram amostradas na análise. Alguns padrões interessantes emergem, como no caso da otimização da conformação ativa que amostrou cinco resíduos de resistência, contrastante a otimização da conformação inativa, que amostrou apenas um resíduo de resistência (V299L). Portanto o método notavelmente capturou que resíduos de resistência são capazes de minimizar a energia da conformação ativa. Além disto, a otimização na forma inativa não alterou a preferência pelo T315, enquanto a otimização da conformação ativa obteve as mutações T315M, que confere resistência, e T315L, que é similar à importante mutação por isoleucina no fenótipo de resistência a inibidores. Este resultado realça a preferência por essa mutação na forma ativa com a presença de outras mutações, sugerindo um efeito cooperativo entre as mutações. Já a otimização da forma Src-like mostrou que a sequência selvagem já é ótima para esta conformação, corroborando o resultado anterior, onde sua energia de conformação é muito menor que as outras duas conformações (**Figura 2.5a**), e sugerindo que o resultado anterior não seria um mero produto do acaso.

Tabela 2.1 Otimização de sequência da Abl1-KD em posições de resistência usando como *input* as conformações estudadas (PDBs 2hyy, 2g1t, 2g2i). Resíduos em itálico são quimicamente similares aos resíduos de resistência; marcação em amarelo indica resíduos de resistência recuperados pela simulação; ciano destaca os resíduos nativos. As dez sequências de menor energia de 750 ciclos de otimização foram analisadas.

WT	mutação de resistência	otimização conf. inativa	otimização conf. Src-like	otimização conf. ativa
M244	M244V	<i>M244I</i>	<i>M244F</i>	<i>M244P</i>
G250	G250E	G250N	WT	G250E, G250N
Q252	Q252H	<i>Q252K</i>	WT	Q252K, Q252D, WT
Y253	Y253H	Y253L	WT	Y253D
E255	E255V, E255M, E255K	<i>E255A</i>	WT	WT, E255Q, E255K
V299	V299L	V299L	WT	V299Q
F311	F311I	<i>F311L</i>	WT	<i>F311M</i>
T315	T315I, T315M	WT	T315M	T315L, T315M
F317	F317L	T317Y	WT	T317M
F359	F359V, F359C	F359K	WT	F359V
H396	H396P, H396R	H396N	H396S	H396P

2.3.3 – Evolução e variabilidade de resíduos da Abl1-KD

A sinalização celular é um mecanismo que orchestra a função celular de acordo com estímulos externos. Em humanos, a sinalização celular é tão importante que as proteínas quinases compõe uma das maiores famílias de proteínas do genoma humano (Kostich et al., 2002). Essa importância é compartilhada com vários organismos multicelulares desde fungos até animais (Manning et al., 2002). Interessantemente, arqueas também possuem certos mecanismos de sinalização celular feito por proteínas quinases (Kennelly, 2014),

destacando ainda mais a evolução e permeação destas famílias de proteínas nos seres vivos.

Com os avanços nas últimas décadas dos métodos computacionais aplicados à teoria evolutiva é possível inferir sequências de proteínas antecessoras às que observamos hoje em dia, o que permite averiguar como as pressões seletivas no passado ditaram as características funcionais das proteínas atuais. Desta forma, buscou-se verificar se ancestrais da Abl1-KD exibem em suas sequências algum padrão evolutivo que influenciaria o aparecimento de mutações causadoras de resistência.

Os aspectos evolutivos da Abl1-KD foram avaliados a partir de uma árvore filogenética construída usando o método de máxima verossimilhança (ML) (**Figura 2.8a**) contendo 41 sequências que se restringem à família de tirosina quinases e os respectivos subgrupos familiares Abl, Src, Tec, Csk (Barker & Dayhoff, 1982); um grupo da proteína quinase A (PKA), que é uma relativa mais distante das quinases anteriores, e um *outgroup* formado pelo domínio quinase de receptores de membrana do fator de crescimento de fibroblastos. Ao longo da árvore filogenética na **Figura 2.8b**, estão os nós que representam eventos de especiação, destacando aqueles que posicionados na linha evolutiva levam à atual sequência da Abl1-KD, também são chamados de ancestrais. Deste modo, 10 sequências ancestrais foram inferidas pelo método de ML, nomeadas de anc0 PhyML (mais recente) para anc9 PhyML (mais distante). Também foi possível observar na árvore filogenética eventos de segregação entre as subfamílias, como por exemplo o ancestral em comum das subfamílias Abl e Src (anc6) e o ancestral em comum de todas as 4 subfamílias presentes (anc8, **Figura 2.8c**). As sequências mais prováveis de cada ancestral foram comparadas em similaridade com a sequência da Abl1-KD atual (círculos vazios, **Figura 2.9a**), e como esperado, a similaridade de resíduos diminui quanto mais distante é o ancestral. Interessantemente, outro estudo, que também determinou os ancestrais da Abl1-KD usando método bayesiano de inferência, reconstruiu e caracterizou um ancestral em comum às enzimas dos grupos Abl e Src (Wilson et al., 2015), o qual é igualmente distante ao Abl1-KD, ~68% em similaridade de sequência, quanto o anc6 PhyML aqui previsto (**Figura 2.9a**), reforçando robustez da nossa análise da evolução desse grupo de quinases. A

explicação das sequências ancestrais PhyML + Rosetta (quadrados preenchidos, **Figura 2.9a**) será introduzida na próxima sessão.

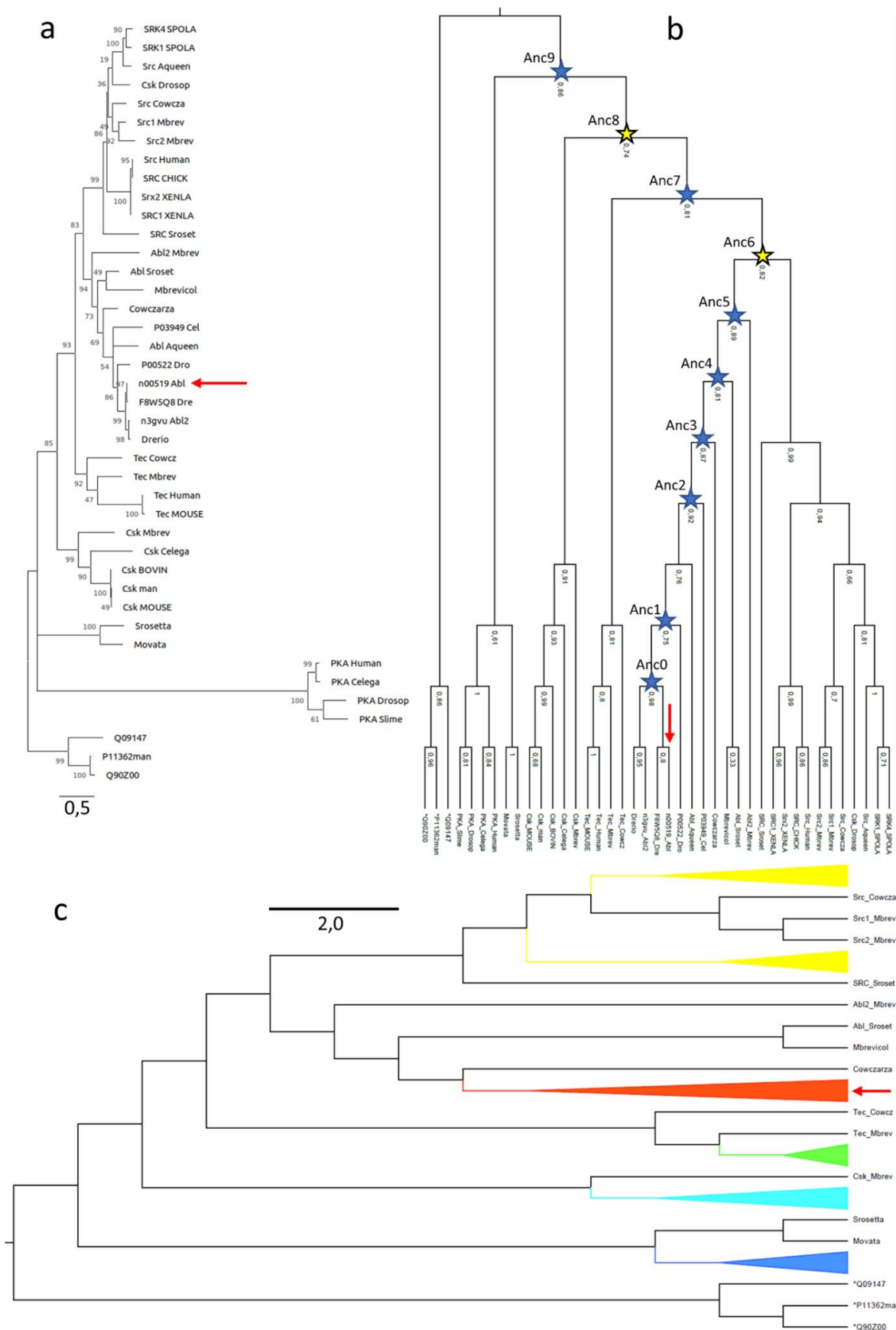


Figura 2.8. Árvore filogenética e ancestrais de Abl1-KD. Estrelas marcam os ancestrais que tiveram suas sequências inferidas e destacada em amarelo os ancestrais que foram experimentalmente avaliados. Setas vermelhas indicam o local da sequência Abl1-KD na árvore. Ramos condensados em amarelo, vermelho, verde, ciano e azul correspondem às subfamílias Src, Abl, Tec, Csk e PKA, respectivamente.

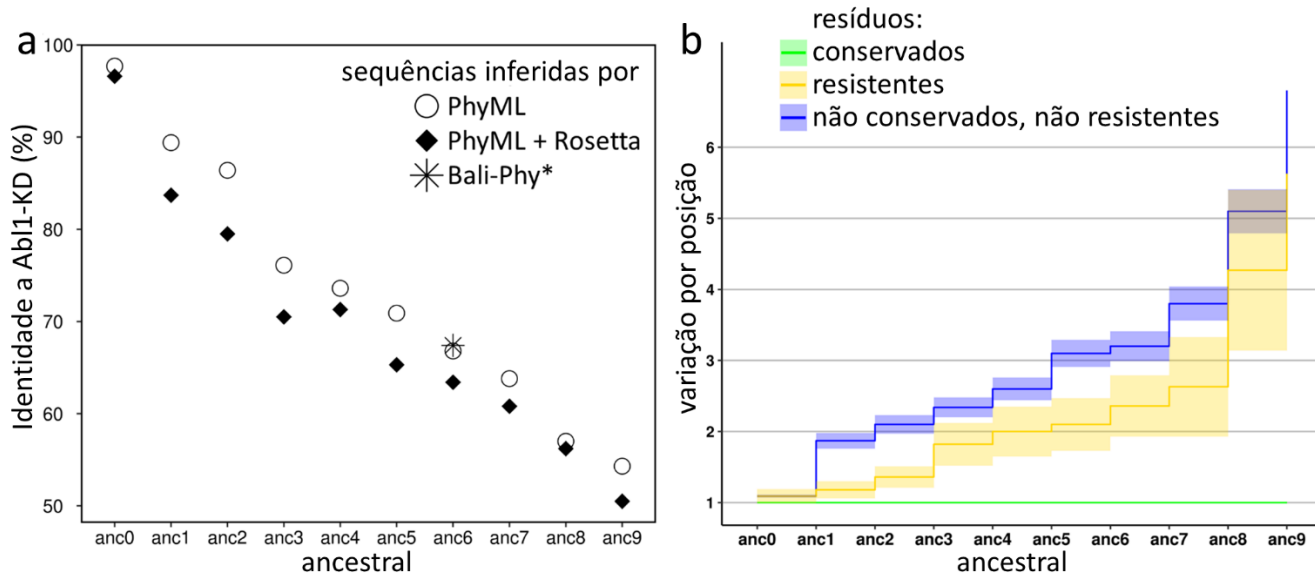


Figura 2.9. a: Comparação de sequências ancestrais com a atual Abl1-KD; PhyML = sequências inferidas apenas por filogenia, PhyML + Rosetta = sequências inferidas por filogenia e submetidas a protocolo do Rosetta, Bali-Phy* = sequência ancestral comum (PDB 4csv) das famílias tirosina quinase Abl1 e Src publicado em outro estudo (Wilson et al., 2015). **b:** variabilidade de resíduos ao longo de sua trajetória evolutiva obtido através do PSSM de cada ancestral.

Finalmente, para avaliar a variação de resíduos das quinases ancestrais de Abl1-KD, as probabilidades de ocorrência dos 20 possíveis resíduos para cada posição na sequência dos ancestrais foram obtidas a partir da árvore filogenética e do alinhamento de sequências (MSA, de *Multiple Sequence Alingment*), extraídas e salvas como PSSM (do inglês *Position-Specific Scoring Matrix*). Esta matriz calcula a frequência do tipo de aminoácido por posição, sendo que quanto maior o *score* do PSSM para determinado aminoácido em certa posição, maior sua probabilidade de ser observado. Na análise de variabilidade foram contadas em cada posição da sequência a quantidade de aminoácido com *score* de probabilidade diferente de zero. Em geral, posições e resíduos que são críticas para a função da proteína tendem a ser mais conservadas, exibindo pouca variação ao longo da evolução. Na **Figura 2.9b** é mostrado que em posições de importante funcionalidade para as quinases, ou seja, os resíduos conservados, a variabilidade

é restrita para apenas um aminoácido (linha verde). Uma análise comparativa interessante envolve a variabilidade das 12 posições que conferem resistência à Abl1-KD (linha amarela) e as posições restantes da sequência (não conservadas e nem ligadas à resistência, linha azul), interpretadas como mutações neutras que foram acumuladas na evolução (Kimura, 1991). No grupo de posições resistentes, a variabilidade de resíduos é menor que no grupo que representa a deriva genética neutra até o anc8 (ancestral em comum das tirosinas quinases), quando então o erro padrão da média (SEM) se sobrepõe à média de variabilidade (**Figura 2.9b**). Em conclusão, os ancestrais mais recentes da Abl1-KD não mostraram muitas substituições em posições ditas de resistência na sequência atual, decerto por conta de riscos associados a um contexto passado, onde trocas poderiam talvez prejudicar a função ou estabilidade das enzimas ancestrais. Mas, no cenário atual, a sobrevivência de clones celulares cancerígenos após o uso de inibidores de quinases é determinada se a mutação em Abl1-KD compensa esses potenciais efeitos negativos com uma menor afinidade à droga inibidora.

2.3.4 – Atividade enzimática dos ancestrais de Abl1-KD e *design* de sequências com informações conformacionais

Até então, a análise de variabilidade de resíduo por posição em sequências ancestrais mostrada anteriormente se limita a um reino de probabilidades. Então buscamos validar a plausibilidade funcional de catálise das sequências ancestrais obtidas na análise filogenética (PhyML), inserindo duas sequências (anc6 PhyML e anc8 PhyML, **Figura 2.9**) em plasmídeo de expressão para serem produzidas na sua forma recombinante em *E. coli* e posteriormente purificadas como mostra a **Figura 2.10**. Na **Figura 2.11** são mostradas as cromatografias de SEC e testes de atividade enzimática, atestando a viabilidade funcional dessas enzimas. Ambas as ancestrais tiveram o mesmo tempo de retenção no SEC que a enzima selvagem atual. Na comparação de atividade, a anc6 PhyML é cerca de 3 vezes mais ativa comparada à selvagem, sugerindo que a Abl1-KD atual teve sua atividade diminuída desde a especiação entre Abl1 e Src. Em outro estudo que primeiramente caracterizou o ancestral inferido por modelo Bayesiano, equivalente ao anc6 PhyML aqui inferido, não mostrou mudanças significativas

na atividade comparada à selvagem (Wilson et al., 2015), indicando que pequenas variações de funcionalidade podem ser esperadas a depender do tipo de algoritmo usado para calcular seqüências ancestrais.

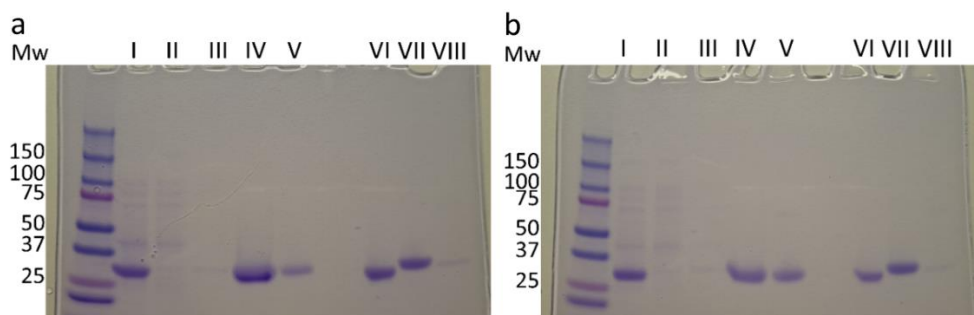


Figura 2.10. Purificação de ancestrais da Abl1-KD. Raias de I a IV referem-se a etapas de purificação das ancestrais. **a:** raia V= anc6 Rosetta, VI = anc6 PhyML, VII = anc8 PhyML, VIII = Abl1-KD atual. **b:** raia V= anc8 Rosetta, VI = anc6 PhyML, VII = anc8 PhyML, VIII = Abl1-KD atual.

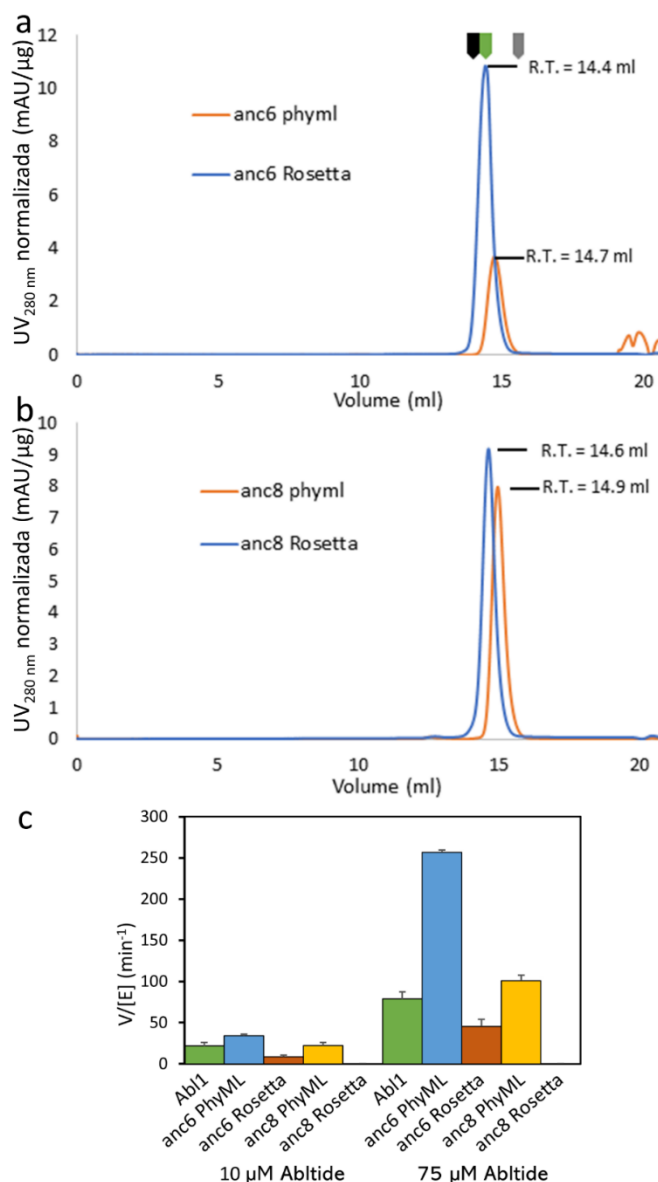


Figura 2.11. Validação experimental de ancestrais da Abl1-KD. **a,b:** SEC de ancestrais inferidas por filogenia apenas em linha laranja e que foram processadas pelo Rosetta em linha azul; marcações a cima dos picos indicam o tempo de retenção da ovoalbumina (preta), Abl1-KD (verde) e anidrase carbônica (cinza). **c:** Atividade quinase de Abl1-KD (verde) e suas ancestrais com o substrato Abltide; dados são médias \pm SD de quatro replicatas. anc8 Rosetta não exibiu alguma atividade.

Com as lições aprendidas das energias de conformação nas sessões anteriores e os resultados filogenéticos de Abl1-KD, questionamos se unificar informações conformacionais com as probabilidades de aminoácidos por posição nas sequências ancestrais levaria a enzimas funcionais. Então foi aplicado um método de *design* nas sequências ancestrais para redefinir as suas sequências inicialmente obtidas usando as restrições conformacionais da forma ativa e inativa, junto com o PSSM dos respectivos ancestrais.

As sequências de ancestrais modificadas após este *design* são chamadas de PhyML + Rosetta e estão representadas na forma de porcentagem de identidade relativa à sequência da Abl1-KD mostrados na **Figura 2.9a** (quadrados preenchidos). Para efeito de comparação, apenas os equivalentes ao anc6 e anc8 foram escolhidas para testes em laboratório. Nas **Figura 2.10 e 2.11** são mostrados as purificações e ensaios com anc6 PhyML + Rosetta e anc8 PhyML + Rosetta. Ambas as ancestrais modificadas foram solúveis e tiveram tempos de retenção similares à selvagem atual Abl1-KD. No que diz respeito as suas atividades enzimáticas, as enzimas desenhadas com o Rosetta foram claramente negativamente afetadas, ao ponto que a anc8 Rosetta se tornou inativa.

Esses resultados mostram que métodos filogenéticos de inferência superam os métodos com base estruturais para capturar a função da proteína, provavelmente porque simulações computacionais de biomoléculas podem ser muito boas em derivar as energias de interação entre os átomos, mas ainda não há um algoritmo que traduza a função da biomolécula para componentes fundamentalmente físicas. Mas vale ressaltar que em um contexto como observado na CML, a menor atividade da quinase é o que se almeja com os tratamentos de inibição, então o método computacional apresentado aqui poderia servir como uma maneira de contrabalancear a funcionalidade nos casos em que tais funções devem ser firmemente reguladas. Portanto, incrementar a estabilização da estrutura de

uma proteína poderia ser uma via para ajustar este grau de regulação intrinsecamente.

2.4 – Conclusões e perspectivas

Neste estudo, estratégias de *design* de sequência por métodos computacionais foram hibridizadas com análise da evolução proteica para entender e talvez antecipar a forma como a enzima Abl1-KD tende a adquirir certos resíduos que promovem resistência ao tratamento da CML.

Avaliando sistematicamente a plasticidade da energia conformacional com todos os possíveis mutantes pontuais, determinamos que mutações individuais são capazes de remodelar a energia de conformações opostas, *i.e.*, ativo e inativo. Esta característica não é exclusiva de mutações que conferem resistência, sugerindo que outros mecanismos atuam em conjunto para reduzir a ligação de inibidores na Abl1-KD. Em geral, foi demonstrado que os métodos computacionais podem até certo grau detectar a ocorrência de mutações de resistência. Mas vale ressaltar que neste caso, foi aplicado o conhecimento clínico prévio dos resíduos e posições que exercem resistência a inibidores. Sem o conhecimento prévio, as possibilidades de mutantes de resistência seriam enormes.

A análise da trajetória evolutiva de Abl1-KD destacou as posições que conferem resistência como moderadamente conservadas, então algum tipo de restrição evolutiva é exercido nestas posições. Métodos que tentam prever mutações capazes de conferir resistência podem ser úteis durante o processo de triagem na criação de novos inibidores. É provável que adicionar uma camada de filtro com esse tipo de informação evolutiva discutida nesta tese ajude na rotina de identificação de inibidores eficazes contra alvos moleculares análogos ao BCR-Abl1. Especialmente em momentos atuais onde a capacidade de resistência a tratamentos por doenças neoplásicas e outras patologias vem sendo uma realidade inconveniente em muitos hospitais.

2.5 – Publicação em jornal científico



Received: 12 January 2022 | Revised: 31 March 2022 | Accepted: 10 May 2022

DOI: 10.1002/pro.4354

FULL-LENGTH PAPER



Evaluation of residue variability in a conformation-specific context and during evolutionary sequence reconstruction narrows drug resistance selection in Abl1 tyrosine kinase

Felipe A. M. Otsuka^{1,2} | Sinisa Bjelic¹

¹Department of Chemistry and Biomedical Sciences, Linnaeus University, Kalmar, Sweden

²Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, São Paulo, Brazil

Correspondence

Sinisa Bjelic, Department of Chemistry and Biomedical Sciences, Linnaeus University, Kalmar, Sweden.
Email: sinisa.bjelic@lnu.se

Funding information

Fundação de Amparo à Pesquisa do Estado de São Paulo, Grant/Award Numbers: 2018/18537-4, 2019/25955-0; Vetenskapsrådet, Grant/Award Number: SNIC 2021/5-401; Conselho Nacional de Desenvolvimento Científico; Coordenação de Aperfeiçoamento de Pessoal de Nível Superior; Swedish Research Council

Review Editor: John Kuriyan

Abstract

Diseases with readily available therapies may eventually prevail against the specific treatment by the acquisition of resistance. The constitutively active Abl1 tyrosine kinase known to cause chronic myeloid leukemia is an example, where patients may experience relapse after small inhibitor drug treatment. Mutations in the Abl1 tyrosine kinase domain (Abl1-KD) are a critical source of resistance and their emergence depends on the conformational states that have been observed experimentally: the inactive state, the active state, and the intermediate inactive state that resembles Src kinase. Understanding how resistant positions and amino acid identities are determined by selection pressure during drug treatment is necessary to improve future drug development or treatment decisions. We carry out in silico site-saturation mutagenesis over the Abl1-KD structure in a conformational context to evaluate the in situ and conformational stability energy upon mutation. Out of the 11 studied resistant positions, we determined that 7 of the resistant mutations favored the active conformation of Abl1-KD with respect to the inactive state. When, instead, the sequence optimization was modeled simultaneously at resistant positions, we recovered five known resistant mutations in the active conformation. These results suggested that the Abl1 resistance mechanism targeted substitutions that favored the active conformation. Further sequence variability, explored by ancestral reconstruction in Abl1-KD, showed that neutral genetic drift, with respect to amino acid variability, was specifically diminished in the resistant positions. Since resistant mutations are susceptible to chance with a certain probability of fixation, combining methodologies outlined here may narrow and limit the available sequence space for resistance to emerge, resulting in more robust therapeutic treatments over time.

KEYWORDS

Abl1, ancestral reconstruction, conformational energy, kinase, resistant positions

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs License](https://creativecommons.org/licenses/by-nc-nd/4.0/), which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. *Protein Science* published by Wiley Periodicals LLC on behalf of The Protein Society.

Referências

- Ali, M. H., & Imperiali, B. (2005). Protein oligomerization: How and why. *Bioorganic & Medicinal Chemistry*, *13*(17), 5013–5020. <https://doi.org/10.1016/j.bmc.2005.05.037>
- Anfinsen, C. B. (1973). Principles that Govern the Folding of Protein Chains. *Science*, *181*(4096), 223–230. <https://doi.org/10.1126/science.181.4096.223>
- Asp, N.-G., Dahlqvist, A., & Koldovský, O. (1969). Human small-intestinal β -galactosidases. Separation and characterization of one lactase and one hetero β -galactosidase. *Biochemical Journal*, *114*(2), 351–359. <https://doi.org/10.1042/bj1140351>
- Badieyan, S., Bevan, D. R., & Zhang, C. (2012). Probing the Active Site Chemistry of β -Glucosidases along the Hydrolysis Reaction Pathway. *Biochemistry*, *51*(44), 8907–8918. <https://doi.org/10.1021/bi300675x>
- Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., Wang, J., Cong, Q., Kinch, L. N., Schaeffer, R. D., Millán, C., Park, H., Adams, C., Glassman, C. R., DeGiovanni, A., Pereira, J. H., Rodrigues, A. v., van Dijk, A. A., Ebrecht, A. C., ... Baker, D. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. *Science*, *373*(6557), 871–876. <https://doi.org/10.1126/science.abj8754>
- Banner, D. W., Bloomer, A. C., Petsko, G. A., Phillips, D. C., Pogson, C. I., Wilson, I. A., Corran, P. H., Furth, A. J., Milman, J. D., Offord, R. E., Priddle, J. D., & Waley, S. G. (1975). Structure of chicken muscle triose phosphate isomerase determined crystallographically at 2.5Å resolution: Using amino acid sequence data. *Nature*. <https://doi.org/10.1038/255609a0>
- Bar-Even, A., Noor, E., Savir, Y., Liebermeister, W., Davidi, D., Tawfik, D. S., & Milo, R. (2011). The Moderately Efficient Enzyme: Evolutionary and Physicochemical Trends Shaping Enzyme Parameters. *Biochemistry*, *50*(21), 4402–4410. <https://doi.org/10.1021/bi2002289>
- Barker, W. C., & Dayhoff, M. O. (1982). Viral src gene products are related to the catalytic chain of mammalian cAMP-dependent protein kinase. *Proceedings of the National Academy of Sciences of the United States of America*, *79*(9 D), 2836–2839. <https://doi.org/10.1073/pnas.79.9.2836>
- Bateman, A., Martin, M. J., Orchard, S., Magrane, M., Agivetova, R., Ahmad, S., Alpi, E., Bowler-Barnett, E. H., Britto, R., Bursteinas, B., Bye-A-Jee, H., Coetzee, R., Cukura, A., Silva, A. da, Denny, P., Dogan, T., Ebenezer, T. G., Fan, J., Castro, L. G., ... Zhang, J. (2021). UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Research*, *49*(D1). <https://doi.org/10.1093/nar/gkaa1100>
- Batoulis, H., Schmidt, T. H., Weber, P., Schloetel, J.-G., Kandt, C., & Lang, T. (2016). Concentration Dependent Ion-Protein Interaction Patterns Underlying Protein Oligomerization Behaviours. *Scientific Reports*, *6*(1), 24131. <https://doi.org/10.1038/srep24131>

- BeMiller, J. N. (2004). Carbohydrates. In *Kirk-Othmer Encyclopedia of Chemical Technology*. John Wiley & Sons, Inc.
<https://doi.org/10.1002/0471238961.0301180202051309.a01.pub2>
- Bhattacharya, A., Prajapati, R., Chatterjee, S., & Mukherjee, T. K. (2014). Concentration-Dependent Reversible Self-Oligomerization of Serum Albumins through Intermolecular β -Sheet Formation. *Langmuir*, *30*(49), 14894–14904.
<https://doi.org/10.1021/la5034959>
- Bjelic, S., Nivón, L. G., Çelebi-Ölçüm, N., Kiss, G., Rosewall, C. F., Lovick, H. M., Ingalls, E. L., Gallaher, J. L., Seetharaman, J., Lew, S., Montelione, G. T., Hunt, J. F., Michael, F. E., Houk, K. N., & Baker, D. (2013). Computational Design of Enone-Binding Proteins with Catalytic Activity for the Morita–Baylis–Hillman Reaction. *ACS Chemical Biology*, *8*(4), 749–757. <https://doi.org/10.1021/cb3006227>
- Bojarová-Fialová, P., & Křen, V. (2007). Enzymatic Approaches to O-Glycoside Introduction: Glycosidases. In *Comprehensive Glycoscience* (Vols. 1–4, pp. 453–487). Elsevier. <https://doi.org/10.1016/B978-044451967-2/00013-1>
- Braun, T. P., Eide, C. A., & Druker, B. J. (2020). Response and Resistance to BCR-ABL1-Targeted Therapies. *Cancer Cell*, *37*(4), 530–542.
<https://doi.org/10.1016/j.ccell.2020.03.006>
- Bray, F., Ren, J.-S., Masuyer, E., & Ferlay, J. (2013). Global estimates of cancer prevalence for 27 sites in the adult population in 2008. *International Journal of Cancer*, *132*(5), 1133–1145. <https://doi.org/10.1002/ijc.27711>
- Brinda, K. V., & Vishveshwara, S. (2005). A Network Representation of Protein Structures: Implications for Protein Stability. *Biophysical Journal*, *89*(6), 4159–4170. <https://doi.org/10.1529/biophysj.105.064485>
- Bruce Alberts, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, & Peter Walter. (2009). *Biologia Molecular da Celula* (5th ed.).
- Chiba, S. (1997). Molecular Mechanism in α -Glucosidase and Glucoamylase. *Bioscience, Biotechnology, and Biochemistry*, *61*(8), 1233–1239.
<https://doi.org/10.1271/bbb.61.1233>
- Csermely, P., Palotai, R., & Nussinov, R. (2010). Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Trends in Biochemical Sciences*, *35*(10), 539–546. <https://doi.org/10.1016/j.tibs.2010.04.009>
- Cummins, M. C., Jacobs, T. M., Teets, F. D., DiMaio, F., Tripathy, A., & Kuhlman, B. (2022). α Fold accurately predicts distinct conformations based on the oligomeric state of a de novo designed protein. *Protein Science*, *31*(7).
<https://doi.org/10.1002/pro.4368>
- Danielli, L., Li, X., Tuller, T., & Daniel, R. (2020). Quantifying the distribution of protein oligomerization degree reflects cellular information capacity. *Scientific Reports*, *10*(1), 17689. <https://doi.org/10.1038/s41598-020-74811-5>
- Davies, G. J., Gloster, T. M., & Henrissat, B. (2005). Recent structural insights into the expanding world of carbohydrate-active enzymes. *Current Opinion in Structural Biology*, *15*(6), 637–645. <https://doi.org/10.1016/j.sbi.2005.10.008>

- de Oliveira, P. S. L., Ferraz, F. A. N., Pena, D. A., Pramio, D. T., Morais, F. A., & Schechtman, D. (2016). Revisiting protein kinase–substrate interactions: Toward therapeutic development. *Science Signaling*, *9*(420).
<https://doi.org/10.1126/scisignal.aad4016>
- Deininger, M., Buchdunger, E., & Druker, B. J. (2005). The development of imatinib as a therapeutic agent for chronic myeloid leukemia. In *Blood*.
<https://doi.org/10.1182/blood-2004-08-3097>
- Dickmanns, A., Ballschmiter, M., Liebl, W., & Ficner, R. (2006). Structure of the novel α -amylase AmyC from *Thermotoga maritima*. *Acta Crystallographica Section D Biological Crystallography*, *62*(3), 262–270.
<https://doi.org/10.1107/S0907444905041363>
- Dill, K. A., & MacCallum, J. L. (2012). The protein-folding problem, 50 years on. In *Science*. <https://doi.org/10.1126/science.1219021>
- Dill, K. A., Ozkan, S. B., Shell, M. S., & Weikl, T. R. (2008). The Protein Folding Problem. *Annual Review of Biophysics*, *37*(1), 289–316.
<https://doi.org/10.1146/annurev.biophys.37.092707.153558>
- Dou, J., Vorobieva, A. A., Sheffler, W., Doyle, L. A., Park, H., Bick, M. J., Mao, B., Foight, G. W., Lee, M. Y., Gagnon, L. A., Carter, L., Sankaran, B., Ovchinnikov, S., Marcos, E., Huang, P.-S., Vaughan, J. C., Stoddard, B. L., & Baker, D. (2018). De novo design of a fluorescence-activating β -barrel. *Nature*, *561*(7724), 485–491.
<https://doi.org/10.1038/s41586-018-0509-0>
- Eswar, N., Eramian, D., Webb, B., Shen, M.-Y., & Sali, A. (2008). Protein structure modeling with MODELLER. *Methods in Molecular Biology (Clifton, N.J.)*, *426*, 145–159. https://doi.org/10.1007/978-1-60327-058-8_8
- Ferreira, C., Capella, A. N., Sitnik, R., & Terra, W. R. (1994). Digestive enzymes in midgut cells, endo-and ectoperitrophic contents, and peritrophic membranes of *Spodoptera frugiperda* (lepidoptera) larvae. *Archives of Insect Biochemistry and Physiology*. <https://doi.org/10.1002/arch.940260406>
- Ferreira, C., & Terra, W. R. (1983). Physical and kinetic properties of a plasma-membrane-bound β -glucosidase (cellobiase) from midgut cells of an insect (*Rhynchosciara americana* larva). *Biochemical Journal*, *213*(1), 43–51.
<https://doi.org/10.1042/bj2130043>
- Fischer, E. (1894). Einfluss der Configuration auf die Wirkung der Enzyme. *Berichte Der Deutschen Chemischen Gesellschaft*, *27*(3), 2985–2993.
<https://doi.org/10.1002/cber.18940270364>
- Fleishman, S. J., Leaver-Fay, A., Corn, J. E., Strauch, E. M., Khare, S. D., Koga, N., Ashworth, J., Murphy, P., Richter, F., Lemmon, G., Meiler, J., & Baker, D. (2011). Rosettascripts: A scripting language interface to the Rosetta Macromolecular modeling suite. *PLoS ONE*, *6*(6). <https://doi.org/10.1371/journal.pone.0020161>
- Florindo, R. N., Souza, V. P., Manzine, L. R., Camilo, C. M., Marana, S. R., Polikarpov, I., & Nascimento, A. S. (2018). Structural and biochemical characterization of a

- GH3 β -glucosidase from the probiotic bacteria *Bifidobacterium adolescentis*. *Biochimie*, *148*, 107–115. <https://doi.org/10.1016/j.biochi.2018.03.007>
- Frauenfelder, H., Chen, G., Berendzen, J., Fenimore, P. W., Jansson, H., McMahon, B. H., Stroe, I. R., Swenson, J., & Young, R. D. (2009). A unified model of protein dynamics. *Proceedings of the National Academy of Sciences of the United States of America*. <https://doi.org/10.1073/pnas.0900336106>
- Goldenzweig, A., Goldsmith, M., Hill, S. E., Gertman, O., Laurino, P., Ashani, Y., Dym, O., Unger, T., Albeck, S., Prilusky, J., Lieberman, R. L., Aharoni, A., Silman, I., Sussman, J. L., Tawfik, D. S., & Fleishman, S. J. (2016). Automated Structure- and Sequence-Based Design of Proteins for High Bacterial Expression and Stability. *Molecular Cell*, *63*(2), 337–346. <https://doi.org/10.1016/j.molcel.2016.06.012>
- Goodsell, D. S., & Olson, A. J. (2000). Structural symmetry and protein function. In *Annual Review of Biophysics and Biomolecular Structure*. <https://doi.org/10.1146/annurev.biophys.29.1.105>
- Greaves, M., & Maley, C. C. (2012). Clonal evolution in cancer. *Nature*, *481*(7381), 306–313. <https://doi.org/10.1038/nature10762>
- Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Systematic Biology*, *59*(3). <https://doi.org/10.1093/sysbio/syq010>
- Gutfreund, H. (1995). Kinetics for the Life Sciences. In *Kinetics for the Life Sciences*. <https://doi.org/10.1017/cbo9780511626203>
- Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of Cancer: The Next Generation. *Cell*, *144*(5), 646–674. <https://doi.org/10.1016/j.cell.2011.02.013>
- Hari Krishna, S. (2002). Developments and trends in enzyme catalysis in nonconventional media. *Biotechnology Advances*, *20*(3–4), 239–267. [https://doi.org/10.1016/S0734-9750\(02\)00019-8](https://doi.org/10.1016/S0734-9750(02)00019-8)
- Henzler-Wildman, K., & Kern, D. (2007). Dynamic personalities of proteins. *Nature*, *450*(7172), 964–972. <https://doi.org/10.1038/nature06522>
- Hochhaus, A., Baccarani, M., Silver, R. T., Schiffer, C., Apperley, J. F., Cervantes, F., Clark, R. E., Cortes, J. E., Deininger, M. W., Guilhot, F., Hjorth-Hansen, H., Hughes, T. P., Janssen, J. J. W. M., Kantarjian, H. M., Kim, D. W., Larson, R. A., Lipton, J. H., Mahon, F. X., Mayer, J., ... Hehlmann, R. (2020). European LeukemiaNet 2020 recommendations for treating chronic myeloid leukemia. In *Leukemia* (Vol. 34, Issue 4, pp. 966–984). Springer Nature. <https://doi.org/10.1038/s41375-020-0776-2>
- Hong, P., Koza, S., & Bouvier, E. S. P. (2012). Size-Exclusion Chromatography for the Analysis of Protein Biotherapeutics and their Aggregates. *Journal of Liquid Chromatography & Related Technologies*, *35*(20), 2923–2950. <https://doi.org/10.1080/10826076.2012.743724>
- Iqbal, Z., Aleem, A., Iqbal, M., Naqvi, M. I., Gill, A., Taj, A. S., Qayyum, A., ur-Rehman, N., Khalid, A. M., Shah, I. H., Khalid, M., Haq, R., Khan, M., Baig, S. M., Jamil, A.,

- Abbas, M. N., Absar, M., Mahmood, A., Rasool, M., & Akhtar, T. (2013). Sensitive Detection of Pre-Existing BCR-ABL Kinase Domain Mutations in CD34+ Cells of Newly Diagnosed Chronic-Phase Chronic Myeloid Leukemia Patients Is Associated with Imatinib Resistance: Implications in the Post-Imatinib Era. *PLoS ONE*, *8*(2), e55717. <https://doi.org/10.1371/journal.pone.0055717>
- Irwin H. Segel. (1993). *Enzyme Kinetics: Behavior and Analysis of Rapid Equilibrium and Steady-State Enzyme Systems*.
- Jeng, W.-Y. Y., Wang, N.-C. C., Lin, M.-H. H., Lin, C.-T. T., Liaw, Y.-C. C., Chang, W.-J. J., Liu, C.-I. I., Liang, P.-H. H., & Wang, A. H.-J. H. J. (2011). Structural and functional analysis of three β -glucosidases from bacterium *Clostridium cellulovorans*, fungus *Trichoderma reesei* and termite *Neotermes koshunensis*. *Journal of Structural Biology*, *173*(1), 46–56. <https://doi.org/10.1016/j.jsb.2010.07.008>
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Židek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., ... Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583–589. <https://doi.org/10.1038/s41586-021-03819-2>
- Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*. <https://doi.org/10.1093/molbev/mst010>
- Kennelly, P. J. (2014). Protein Ser/Thr/Tyr phosphorylation in the Archaea. *The Journal of Biological Chemistry*, *289*(14), 9480–9487. <https://doi.org/10.1074/jbc.R113.529412>
- Ketudat Cairns, J. R., & Esen, A. (2010). β -Glucosidases. *Cellular and Molecular Life Sciences*, *67*(20), 3389–3405. <https://doi.org/10.1007/s00018-010-0399-2>
- Khatib, F., Cooper, S., Tyka, M. D., Xu, K., Makedon, I., Popović, Z., Baker, D., & Players, F. (2011). Algorithm discovery by protein folding game players. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(47). <https://doi.org/10.1073/pnas.1115898108>
- Khersonsky, O., Lipsh, R., Avizemer, Z., Ashani, Y., Goldsmith, M., Leader, H., Dym, O., Rogotner, S., Trudeau, D. L., Prilusky, J., Amengual-Rigo, P., Guallar, V., Tawfik, D. S., & Fleishman, S. J. (2018). Automated Design of Efficient and Functionally Diverse Enzyme Repertoires. *Molecular Cell*, *72*(1), 178-186.e5. <https://doi.org/10.1016/j.molcel.2018.08.033>
- KIM, S., KIM, Y., HEGERL, R., CYRKLAFF, M., & KIM, I. (2005). Novel type of enzyme multimerization enhances substrate affinity of oat β -glucosidase. *Journal of Structural Biology*, *150*(1), 1–10. <https://doi.org/10.1016/j.jsb.2004.07.007>
- Kim, S.-Y., & Kim, I.-S. (2004). Oligomeric Structure of β -Glucosidases. *Journal of Photoscience*, *11*(3), 121–127. <http://www.photos.or.kr/journal/issues/11-3-121.pdf>

- Kimura, M. (1991). The neutral theory of molecular evolution: A review of recent evidence. *The Japanese Journal of Genetics*, *66*(4), 367–386. <https://doi.org/10.1266/jjg.66.367>
- Knowles, J. R. (1991). Enzyme catalysis: not different, just better. *Nature*, *350*(6314), 121–124. <https://doi.org/10.1038/350121a0>
- Kostich, M., English, J., Madison, V., Gheyas, F., Wang, L., Qiu, P., Greene, J., & Laz, T. M. (2002). Human members of the eukaryotic protein kinase family. *Genome Biology*, *3*(9), research0043.1. <https://doi.org/10.1186/gb-2002-3-9-research0043>
- Kötzler, M. P., Hancock, S. M., & Withers, S. G. (2014). Glycosidases: Functions, Families and Folds. In *eLS*. Wiley. <https://doi.org/10.1002/9780470015902.a0020548.pub2>
- Krissinel, E., & Henrick, K. (2007). Inference of Macromolecular Assemblies from Crystalline State. *Journal of Molecular Biology*, *372*(3), 774–797. <https://doi.org/10.1016/j.jmb.2007.05.022>
- Kuhlman, B., & Bradley, P. (2019). Advances in protein structure prediction and design. *Nature Reviews Molecular Cell Biology*, *20*(11), 681–697. <https://doi.org/10.1038/s41580-019-0163-x>
- Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*. <https://doi.org/10.1093/molbev/msy096>
- Kumari, N., & Yadav, S. (2019). Modulation of protein oligomerization: An overview. *Progress in Biophysics and Molecular Biology*, *149*, 99–113. <https://doi.org/10.1016/j.pbiomolbio.2019.03.003>
- LAEMMLI, U. K. (1970). Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature*, *227*(5259), 680–685. <https://doi.org/10.1038/227680a0>
- Leaver-Fay, A., Tyka, M., Lewis, S. M., Lange, O. F., Thompson, J., Jacak, R., Kaufman, K., Douglas Renfrew, P., Smith, C. A., Sheffler, W., Davis, I. W., Cooper, S., Treuille, A., Mandell, D. J., Richter, F., Andrew Ban, Y.-E., Fleishman, S. J., Corn, J. E., Kim, D. E., ... Bradley, P. (2011). ROSETTA3: An Object-Oriented Software Suite for the Simulation and Design of Macromolecules. *Methods Enzymol*, *487*, 545–574. <https://doi.org/10.1016/B978-0-12-381270-4.00019-6>
- Li, Y., Zhang, R., Wang, C., Forouhar, F., Clarke, O. B., Vorobiev, S., Singh, S., Montelione, G. T., Szyperski, T., Xu, Y., & Hunt, J. F. (2022). Oligomeric interactions maintain active-site structure in a noncooperative enzyme family. *The EMBO Journal*, *41*(17). <https://doi.org/10.15252/emj.2021108368>
- Lichtenthaler, F. W., & Peters, S. (2004). Carbohydrates as green raw materials for the chemical industry. *Comptes Rendus Chimie*, *7*(2), 65–90. <https://doi.org/10.1016/j.crci.2004.02.002>
- Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., & Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research*, *42*(D1), D490–D495. <https://doi.org/10.1093/nar/gkt1178>

- Lovera, S., Morando, M., Pucheta-Martinez, E., Martinez-Torrecuadrada, J. L., Saladino, G., & Gervasio, F. L. (2015). Towards a Molecular Understanding of the Link between Imatinib Resistance and Kinase Conformational Dynamics. *PLoS Computational Biology*, *11*(11). <https://doi.org/10.1371/journal.pcbi.1004578>
- Ly, H. D., & Withers, S. G. (1999). Mutagenesis of Glycosidases. *Annual Review of Biochemistry*, *68*(1), 487–522. <https://doi.org/10.1146/annurev.biochem.68.1.487>
- Mahatabuddin, S., Hanada, Y., Nishimiya, Y., Miura, A., Kondo, H., Davies, P. L., & Tsuda, S. (2017). Concentration-dependent oligomerization of an alpha-helical antifreeze polypeptide makes it hyperactive. *Scientific Reports*, *7*(1), 42501. <https://doi.org/10.1038/srep42501>
- Manning, G., Plowman, G. D., Hunter, T., & Sudarsanam, S. (2002). Evolution of protein kinase signaling from yeast to man. In *Trends in Biochemical Sciences* (Vol. 27, Issue 10, pp. 514–520). [https://doi.org/10.1016/S0968-0004\(02\)02179-5](https://doi.org/10.1016/S0968-0004(02)02179-5)
- Marana, S. R., Jacobs-Lorena, M., Terra, W. R., & Ferreira, C. (2001). Amino acid residues involved in substrate binding and catalysis in an insect digestive β -glycosidase. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology*, *1545*(1–2), 41–52. [https://doi.org/10.1016/S0167-4838\(00\)00260-0](https://doi.org/10.1016/S0167-4838(00)00260-0)
- Marana, S. R., Terra, W. R., & Ferreira, C. (2000). Purification and properties of a β -glycosidase purified from midgut cells of *Spodoptera frugiperda* (Lepidoptera) larvae. *Insect Biochemistry and Molecular Biology*, *30*(12), 1139–1146. [https://doi.org/10.1016/S0965-1748\(00\)00090-4](https://doi.org/10.1016/S0965-1748(00)00090-4)
- Marana, S. R., Terra, W. R., & Ferreira, C. (2002). The role of amino-acid residues Q39 and E451 in the determination of substrate specificity of the *Spodoptera frugiperda* β -glycosidase. *European Journal of Biochemistry*, *269*(15), 3705–3714. <https://doi.org/10.1046/j.1432-1033.2002.03061.x>
- Mendonça, L. M. F., & Marana, S. R. (2008). The role in the substrate specificity and catalysis of residues forming the substrate aglycone-binding site of a β -glycosidase. *FEBS Journal*, *275*(10), 2536–2547. <https://doi.org/10.1111/j.1742-4658.2008.06402.x>
- Mendonça, L. M. F., & Marana, S. R. (2011). Single mutations outside the active site affect the substrate specificity in a β -glycosidase. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, *1814*(12), 1616–1623. <https://doi.org/10.1016/J.BBAPAP.2011.08.012>
- Merrikkh, H., & Kohli, R. M. (2020). Targeting evolution to inhibit antibiotic resistance. *The FEBS Journal*, *287*(20), 4341–4353. <https://doi.org/10.1111/febs.15370>
- Nakamura, T., Fahmi, M., Tanaka, J., Seki, K., Kubota, Y., & Ito, M. (2019). Genome-Wide Analysis of Whole Human Glycoside Hydrolases by Data-Driven Analysis in Silico. *International Journal of Molecular Sciences*, *20*(24), 6290. <https://doi.org/10.3390/ijms20246290>
- Narayan, B., Fathizadeh, A., Templeton, C., He, P., Arasteh, S., Elber, R., Buchete, N.-V., & Levy, R. M. (2020). The transition between active and inactive conformations of Abl kinase studied by rock climbing and Milestoning. *Biochimica et Biophysica*

- Acta (BBA) - General Subjects*, 1864(4), 129508.
<https://doi.org/10.1016/j.bbagen.2019.129508>
- Noor, S., Taylor, M. C., Russell, R. J., Jermiin, L. S., Jackson, C. J., Oakeshott, J. G., & Scott, C. (2012). Intramolecular Epistasis and the Evolution of a New Enzymatic Function. *PLoS ONE*, 7(6), e39822. <https://doi.org/10.1371/journal.pone.0039822>
- Nowell, P. C. (1976). The Clonal Evolution of Tumor Cell Populations. *Science*, 194(4260), 23–28. <https://doi.org/10.1126/science.959840>
- O. Silva, R., R. de Freitas Filho, J., & C. Rufino Freitas, J. (2018). D-Glucose, a Fascinating Biomolecule: History, Properties, Production and Application. *Revista Virtual de Química*, 10(4), 875–891. <https://doi.org/10.21577/1984-6835.20180064>
- O'hare, T., Zabriskie, M. S., Eiring, A. M., & Deininger, M. W. (2012). Pushing the limits of targeted therapy in chronic myeloid leukaemia. In *Nature Reviews Cancer*. <https://doi.org/10.1038/nrc3317>
- Oruganti, B., & Friedman, R. (2021). Activation of Abl1 Kinase Explored Using Well-Tempered Metadynamics Simulations on an Essential Dynamics Sampled Path. *Journal of Chemical Theory and Computation*, 17(11), 7260–7270. <https://doi.org/10.1021/acs.jctc.1c00505>
- Otsuka, F. A. M., & Bjelic, S. (2022). Evaluation of residue variability in a conformation-specific context and during evolutionary sequence reconstruction narrows drug resistance selection in Abl1 tyrosine kinase. *Protein Science*, 31(7), 1–14. <https://doi.org/10.1002/pro.4354>
- Otsuka, F. A. M., Chagas, R. S., Almeida, V. M., & Marana, S. R. (2020). Homodimerization of a glycoside hydrolase family GH1 β -glucosidase suggests distinct activity of enzyme different states. *Protein Science: A Publication of the Protein Society*, 29(9), 1879–1889. <https://doi.org/10.1002/pro.3908>
- Pauling, L. (1935). The Oxygen Equilibrium of Hemoglobin and Its Structural Interpretation. *Proceedings of the National Academy of Sciences*, 21(4), 186–191. <https://doi.org/10.1073/pnas.21.4.186>
- Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera - A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13), 1605–1612. <https://doi.org/10.1002/jcc.20084>
- Pikis, A., Immel, S., Robrish, S. A., & Thompson, J. (2002). Metabolism of sucrose and its five isomers by *Fusobacterium mortiferum*. *Microbiology*, 148(3), 843–852. <https://doi.org/10.1099/00221287-148-3-843>
- Ren, R. (2005). Mechanisms of BCR-ABL in the pathogenesis of chronic myelogenous leukaemia. In *Nature Reviews Cancer* (Vol. 5, Issue 3, pp. 172–183). <https://doi.org/10.1038/nrc1567>
- Robinson, P. K. (2015). Enzymes: principles and biotechnological applications. *Essays in Biochemistry*, 59, 1–41. <https://doi.org/10.1042/bse0590001>

- Santos, C. R., Tonoli, C. C. C., Trindade, D. M., Betzel, C., Takata, H., Kuriki, T., Kanai, T., Imanaka, T., Arni, R. K., & Murakami, M. T. (2011). Structural basis for branching-enzyme activity of glycoside hydrolase family 57: Structure and stability studies of a novel branching enzyme from the hyperthermophilic archaeon *Thermococcus Kodakaraensis* KOD1. *Proteins: Structure, Function, and Bioinformatics*, *79*(2), 547–557. <https://doi.org/10.1002/prot.22902>
- Schmitt, M. W., Pritchard, J. R., Leighow, S. M., Aminov, B. I., Beppu, L., Kim, D. S., Hodgson, J. G., Rivera, V. M., Loeb, L. A., & Radich, J. P. (2018). Single-Molecule Sequencing Reveals Patterns of Preexisting Drug Resistance That Suggest Treatment Strategies in Philadelphia-Positive Leukemias. *Clinical Cancer Research*, *24*(21), 5321–5334. <https://doi.org/10.1158/1078-0432.CCR-18-0167>
- Seeliger, M. A., Young, M., Henderson, M. N., Pellicena, P., King, D. S., Falick, A. M., & Kuriyan, J. (2005). High yield bacterial expression of active c-Abl and c-Src tyrosine kinases. *Protein Science*, *14*(12), 3135–3139. <https://doi.org/10.1110/ps.051750905>
- Sela, I., Ashkenazy, H., Katoh, K., & Pupko, T. (2015). GUIDANCE2: Accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters. *Nucleic Acids Research*, *43*(W1). <https://doi.org/10.1093/nar/gkv318>
- Seshadri, S., Akiyama, T., Opassiri, R., Kuaprasert, B., & Cairns, J. K. (2009). Structural and Enzymatic Characterization of Os3BGlu6, a Rice β -Glucosidase Hydrolyzing Hydrophobic Glycosides and (1 \rightarrow 3)- and (1 \rightarrow 2)-Linked Disaccharides. *Plant Physiology*, *151*(1), 47–58. <https://doi.org/10.1104/pp.109.139436>
- Shah, N. P., & Sawyers, C. L. (2003). Mechanisms of resistance to STI571 in Philadelphia chromosome-associated leukemias. *Oncogene*, *22*(47 REV. ISS. 6), 7389–7395. <https://doi.org/10.1038/sj.onc.1206942>
- Shtivelman, E., Lifshitz, B., Gale, R. P., Roe, B. A., & Canaani, E. (1986). Alternative splicing of RNAs transcribed from the human abl gene and from the bcr-abl fused gene. *Cell*, *47*(2), 277–284. [https://doi.org/10.1016/0092-8674\(86\)90450-2](https://doi.org/10.1016/0092-8674(86)90450-2)
- Shukla, A., & Guptasarma, P. (2004). Folding of β/α -Unit Scrambled Forms of *S. cerevisiae* Triosephosphate Isomerase: Evidence for Autonomy of Substructure Formation and Plasticity of Hydrophobic and Hydrogen Bonding Interactions in Core of $(\beta/\alpha)_8$ -Barrel. *Proteins: Structure, Function and Genetics*. <https://doi.org/10.1002/prot.20066>
- Singh, P., Vandemeulebroucke, A., Li, J., Schulenburg, C., Fortunato, G., Kohen, A., Hilvert, D., & Cheatum, C. M. (2021). Evolution of the Chemical Step in Enzyme Catalysis. *ACS Catalysis*, *11*(11), 6726–6732. <https://doi.org/10.1021/acscatal.1c00442>
- Smith, P. K., Krohn, R. I., Hermanson, G. T., Mallia, A. K., Gartner, F. H., Provenzano, M. D., Fujimoto, E. K., Goeke, N. M., Olson, B. J., & Klenk, D. C. (1985). Measurement of protein using bicinchoninic acid. *Analytical Biochemistry*. [https://doi.org/10.1016/0003-2697\(85\)90442-7](https://doi.org/10.1016/0003-2697(85)90442-7)

- Southall, N. T., Dill, K. A., & Haymet, A. D. J. (2002). A View of the Hydrophobic Effect. *The Journal of Physical Chemistry B*, *106*(3), 521–533. <https://doi.org/10.1021/jp015514e>
- Souza, V. P. (2017). *Papel das Redes Estruturais Proteicas nas Propriedades de uma Beta-Glicosidase*. São Paulo University.
- Souza, V. P., Ikegami, C. M., Arantes, G. M., & Marana, S. R. (2016). Protein thermal denaturation is modulated by central residues in the protein structure network. *FEBS Journal*, *283*(6), 1124–1138. <https://doi.org/10.1111/febs.13659>
- Soverini, S., Mancini, M., Bavaro, L., Cavo, M., & Martinelli, G. (2018). Chronic myeloid leukemia: the paradigm of targeting oncogenic tyrosine kinase signaling and counteracting resistance for successful cancer therapy. *Molecular Cancer*, *17*(1), 49. <https://doi.org/10.1186/s12943-018-0780-6>
- Tamaki, F. K., Souza, D. P., Souza, V. P., Ikegami, C. M., Farah, C. S., & Marana, S. R. (2016). Using the Amino Acid Network to Modulate the Hydrolytic Activity of β -Glycosidases. *PLOS ONE*, *11*(12), e0167978. <https://doi.org/10.1371/journal.pone.0167978>
- Taylor, S. S., Keshwani, M. M., Steichen, J. M., & Kornev, A. P. (2012). Evolution of the eukaryotic protein kinases as dynamic molecular switches. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1602), 2517–2528. <https://doi.org/10.1098/rstb.2012.0054>
- Terra, W. R., Ferreira, C., & de Bianchi, A. G. (1979). Distribution of digestive enzymes among the endo- and ectoperitrophic spaces and midgut cells of *Rhynchosciara* and its physiological significance. *Journal of Insect Physiology*, *25*(6), 487–494. [https://doi.org/10.1016/S0022-1910\(79\)80007-4](https://doi.org/10.1016/S0022-1910(79)80007-4)
- Teufl, M., Zajc, C. U., & Traxlmayr, M. W. (2022). Engineering Strategies to Overcome the Stability–Function Trade-Off in Proteins. *ACS Synthetic Biology*, *11*(3), 1030–1039. <https://doi.org/10.1021/acssynbio.1c00512>
- Thorlby, G., Fourrier, N., & Warren, G. (2004). The *SENSITIVE TO FREEZING2* Gene, Required for Freezing Tolerance in *Arabidopsis thaliana*, Encodes a β -Glucosidase. *The Plant Cell*, *16*(8), 2192–2203. <https://doi.org/10.1105/tpc.104.024018>
- Tokuriki, N., Stricher, F., Serrano, L., & Tawfik, D. S. (2008). How Protein Stability and New Functions Trade Off. *PLoS Computational Biology*, *4*(2), e1000002. <https://doi.org/10.1371/journal.pcbi.1000002>
- Trott, O., & Olson, A. J. (2009). AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, NA-NA. <https://doi.org/10.1002/jcc.21334>
- Trudeau, D. L., & Tawfik, D. S. (2019). Protein engineers turned evolutionists—the quest for the optimal starting point. *Current Opinion in Biotechnology*, *60*, 46–52. <https://doi.org/10.1016/j.copbio.2018.12.002>
- Umar, A., Dunn, B. K., & Greenwald, P. (2012). Future directions in cancer prevention. *Nature Reviews Cancer*, *12*(12), 835–848. <https://doi.org/10.1038/nrc3397>

- Vasan, N., Baselga, J., & Hyman, D. M. (2019). A view on drug resistance in cancer. *Nature*, *575*(7782), 299–309. <https://doi.org/10.1038/s41586-019-1730-1>
- Vogt, A. D., & di Cera, E. (2012). Conformational selection or induced fit? A critical appraisal of the kinetic mechanism. *Biochemistry*, *51*(30), 5894–5902. <https://doi.org/10.1021/bi3006913>
- Weikl, T. R., & Paul, F. (2014). Conformational selection in protein binding and function. *Protein Science*, *23*(11), 1508–1518. <https://doi.org/10.1002/pro.2539>
- Wilson, C., Agafonov, R. v., Hoemberger, M., Kutter, S., Zorba, A., Halpin, J., Buosi, V., Otten, R., Waterman, D., Theobald, D. L., & Kern, D. (2015). Using ancient protein kinases to unravel a modern cancer drug's mechanism. In *Science* (Vol. 347, Issue 6224, pp. 882–886). American Association for the Advancement of Science. <https://doi.org/10.1126/science.aaa1823>
- Withers, S. (2001). Mechanisms of glycosyl transferases and hydrolases. *Carbohydrate Polymers*, *44*(4), 325–337. [https://doi.org/10.1016/S0144-8617\(00\)00249-6](https://doi.org/10.1016/S0144-8617(00)00249-6)
- Wolf-Watz, M., Thai, V., Henzler-Wildman, K., Hadjipavlou, G., Eisenmesser, E. Z., & Kern, D. (2004). Linkage between dynamics and catalysis in a thermophilic-mesophilic enzyme pair. *Nature Structural & Molecular Biology*, *11*(10), 945–949. <https://doi.org/10.1038/nsmb821>
- Zabriskie, M. S., Eide, C. A., Tantravahi, S. K., Vellore, N. A., Estrada, J., Nicolini, F. E., Khoury, H. J., Larson, R. A., Konopleva, M., Cortes, J. E., Kantarjian, H., Jabbour, E. J., Kornblau, S. M., Lipton, J. H., Rea, D., Stenke, L., Barbany, G., Lange, T., Hernández-Boluda, J.-C., ... O'Hare, T. (2014). BCR-ABL1 Compound Mutations Combining Key Kinase Domain Positions Confer Clinical Resistance to Ponatinib in Ph Chromosome-Positive Leukemia. *Cancer Cell*, *26*(3), 428–442. <https://doi.org/10.1016/j.ccr.2014.07.006>
- Zanphorlin, L. M., de Giuseppe, P. O., Honorato, R. V., Tonoli, C. C. C., Fattori, J., Crespim, E., de Oliveira, P. S. L., Ruller, R., & Murakami, M. T. (2016). Oligomerization as a strategy for cold adaptation: Structure and dynamics of the GH1 β -glucosidase from *Exiguobacterium antarcticum* B7. *Scientific Reports*, *6*(1), 23776. <https://doi.org/10.1038/srep23776>
- Zhou, W., Šmidlehner, T., & Jerala, R. (2020). Synthetic biology principles for the design of protein with novel structures and functions. *FEBS Letters*, *594*(14), 2199–2212. <https://doi.org/10.1002/1873-3468.13796>

Súmula curricular

1. Dados Pessoais

Nome: Felipe Akihiro Melo Otsuka

Local e data de nascimento: Aracaju/SE, 05/03/1993

2. Educação

Colégio Arquidiocesano, Aracaju/SE – Brasil, 2008 – 2010

Universidade Federal de Sergipe, São Cristovão/SE – Brasil, 2011 – 2017, graduação em Ciências Farmacêuticas.

Arizona State University, Tempe/AZ – USA, 2014 – 2015, graduação sanduiche em Bioquímica.

Instituto de Química, Universidade de São Paulo, São Paulo/SP – Brasil, 2017 – 2023, doutorado em Ciências Biológicas (Bioquímica).

3. Formação Complementar

Extensão universitária em IX Curso de Verão em Bioquímica e Biologia Molecular. Universidade de São Paulo, São Paulo/SP – Brasil, 2014.

MBA em *Data Science and Analytics*. USP/ESALQ, São Paulo/SP – Brasil, 2022 – 2023.

4. Ocupação

Doutorado em Ciências Biológicas (Bioquímicas) Instituto de Química, Universidade de São Paulo, São Paulo/SP, 2017 – 2023.

5. Publicações

Otsuka, Felipe A.M.; Santos, Rodrigo B.; Chaves, Larissa F.; Santos, Rosangela S.; Chaves Filho, Adriano B.; Miyamoto, Sayuri; Matos, Humberto R. Identification of caffeic acid and rutin by UHPLC MS/MS and antioxidant

activity of *Commelina erecta* Lineu. in cell culture. *Anais da Academia Brasileira de Ciências*, v. 92, p. 1/e20190491-10, 2020.

Otsuka, Felipe A. M.; Chagas, Rafael S.; Almeida, Vitor M.; Marana, Sandro R. Homodimerization of a glycoside hydrolase family GH1 β -glucosidase suggests distinct activity of enzyme different states. *Protein Science*, v. 29, p. 1879-1889, 2020.

Otsuka, Felipe A. M.; Bjelic, Sinisa. Evaluation of residue variability in a conformation-specific context and during evolutionary sequence reconstruction narrows drug resistance selection in Abl1 tyrosine kinase. *Protein Science*, v. 31, p. e4354, 2022.