



UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE CIÊNCIAS
CURSO DE GRADUAÇÃO EM BIOTECNOLOGIA

IGOR OLIVEIRA DUARTE

**ANÁLISE DA VARIAÇÃO DE ESTABILIDADE CONFORMACIONAL DE
PROTEÍNAS DECORRENTE DE MUTAÇÕES PONTUAIS**

FORTALEZA

2017

IGOR OLIVEIRA DUARTE

ANÁLISE DA VARIAÇÃO DE ESTABILIDADE CONFORMACIONAL DE PROTEÍNAS
DECORRENTE DE MUTAÇÕES PONTUAIS

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Biotecnologia do Centro de Ciências da Universidade Federal do Ceará, como requisito parcial à obtenção do grau de bacharel em Biotecnologia.

Orientador: Dr. Raphael Trevizani

FORTALEZA

2017

IGOR OLIVEIRA DUARTE

ANÁLISE DA VARIAÇÃO DE ESTABILIDADE CONFORMACIONAL DE PROTEÍNAS
DECORRENTE DE MUTAÇÕES PONTUAIS

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Biotecnologia do Centro de Ciências da Universidade Federal do Ceará, como requisito parcial à obtenção do grau de bacharel em Biotecnologia.

Aprovada em: 27 de junho de 2017

BANCA EXAMINADORA

Dr. Raphael Trevizani (Orientador)
Fundação Oswaldo Cruz (Fiocruz-CE)

Prof. Dr. André Luis Coelho da Silva
Universidade Federal do Ceará (UFC)

Dr. Marcos Roberto Lourenzoni
Fundação Oswaldo Cruz (Fiocruz-CE)

À minha família, por sua capacidade de acreditar
e investir em mim.

AGRADECIMENTOS

Aos meu pais, **Cléa** e **Neto**, que fizeram o possível para me proporcionar a melhor formação e garantir que não me faltasse nada. Ao meu irmão **Yuri**, por me ajudar a tomar as decisões mais difíceis e me apoiar em todos os momentos, ao meu irmão **Iago**, pela companhia e confiança.

À **Fundação Oswaldo Cruz/Ceará** pelos recursos cedidos.

Ao meu orientador **Raphael Trevizani** pela paciência em me ensinar e pelo estímulo a aprimorar meus conhecimentos.

Ao professor **André Luís Coelho**, por ter me orientado por metade da minha graduação, proporcionando-me espaço e auxílio que foram essenciais para a minha formação, e por participar da minha banca examinadora.

Aos professores participantes da banca examinadora **Marcos Roberto Lourenzoni** e suplente **Márjory Lima Holanda** pelo tempo e pelas valiosas colaborações e sugestões.

A todos os **professores do curso de Biotecnologia da UFC**, pelos conhecimentos passados e por estarem sempre dispostos a ajudar os alunos em sua formação. À **Profa. Dirce Fernandes de Melo** e a **Gilmar Ferreira da Costa** pela disponibilidade integral em auxiliar os alunos do curso e pela confiança.

À minha colega de trabalho **Débora Albuquerque**, pelas trocas de conhecimento, pelos momentos descontraídos e pela disposição em ajudar sempre que precisei.

Aos meus amigos de curso e da vida: **Geovanny Moreno**, por estar sempre comigo, **Carol Xavier**, por sempre me ensinar alguma coisa sobre a vida, **Fabrcia Diniz**, pela companhia inestimável, **Augusto Filho**, pelos conselhos e apoio incondicional, **Tayná Fiúza**, por eu sempre poder contar com sua amizade e cumplicidade, **Mateus Pinheiro**, por me salvar sempre que sinto que estou me afogando, e **Marcus Rafael**, por acreditar em mim até quando eu não pude.

Aos meus colegas de curso, de trabalho e da vida, por me ajudarem a sempre continuar seguindo meu caminho.

“Make good **art.**”

(Neil Gaiman)

RESUMO

Proteínas possuem diversas aplicações biotecnológicas e são particularmente convenientes devido a sua especificidade. Porém, limitações naturais de proteínas ainda são uma barreira para sua utilização industrial, como na falta de estabilidade em temperaturas altas, afinidade insuficiente pelo substrato e baixa eficiência catalítica; no caso da indústria farmacêutica, proteínas têm seu uso restrito por conta de seu potencial imunogênico. A engenharia de proteínas é capaz de alterar características de proteínas através de mutações em sua estrutura primária, porém, em qualquer mutação conduzida em proteínas, a estabilidade é uma característica que deve ser preservada. Mutagênese sítio-dirigida é a maneira mais eficiente de geração de mutantes, entretanto, métodos computacionais ainda não são capazes de prever corretamente como mutações devem ser preferencialmente conduzidas. Este trabalho teve como objetivo estudar a maneira como descritores que envolvem estrutura do sítio de mutação e identidade de aminoácidos refletem na variação de estabilidade após mutação. A investigação foi baseada em um banco de dados experimental que reúne 3605 mutações pontuais e sua variação de estabilidade calculada em termos de $\Delta\Delta G$, que foi o parâmetro utilizado para diferenciar mutações com impacto neutro na estabilidade ($-1 \text{ kcal/mol} < \Delta\Delta G < 1 \text{ kcal/mol}$), de mutações com impacto favorável ($\Delta\Delta G \leq -1 \text{ kcal/mol}$). Dois descritores estruturais foram avaliados para caracterizar a influência do sítio de mutação: área superficial acessível ao solvente (SASA) e estrutura secundária (SS); e três descritores de identidade foram analisados para caracterizar os resíduos envolvidos na substituição: volume, massa molecular e hidrofobicidade. Os resultados apontam que a SASA possui influência direta na estabilidade, já que o valor de SASA está relacionado à proximidade de um aminoácido ao centro hidrofóbico. Estrutura secundária, porém, não apresentou grande importância em diferenciar mutações neutras de favoráveis. Massa molar e volume foram considerados equivalentes e apresentaram influência na estabilidade, ao contrário da hidrofobicidade, que apresentou baixa relevância na distinção de mutações neutras de favoráveis. Os resultados esclarecem algumas das relações entre os descritores e a estabilidade, porém a análise mais aprofundada dos descritores e a investigação de novos descritores ainda são necessárias a fim de que os melhores parâmetros para uma mutação sejam preditos.

Palavras-chave: Predição de estabilidade. Mutação sítio-dirigida. Análise estatística.

ABSTRACT

Proteins have a number of biotechnological applications and are particularly convenient for their specificity. However, the industrial use of proteins can be limited by their lack of stability in high temperatures, insufficient affinity by the substrate and the number of reactions it catalyzes, whereas the therapeutic use is restricted by their potential immunogenicity. A protein can be engineered for new purposes, such as increased affinity, via a series of mutations in its primary sequence, but all protein engineering projects must account for the protein stability. Site-directed mutagenesis is the cheapest way of generating new mutants, however, *in silico* methods are still inefficient in predicting how it should be preferably conducted. This work aimed to study how mutation site and amino acid features reflect on the change of stability of a protein structure that undergoes a single-point mutation. The investigation was based on an experimental dataset of 3605 single-point mutants along with their $\Delta\Delta G$ values, which were used to classify mutations in two groups of interest: neutral mutations, when there is no significant effect on stability, or favorable ones, when $\Delta\Delta G$ is smaller than -1 kcal/mol. Two structural features were studied to account for choosing the most suitable mutation site for each group of interest: solvent accessibility surface area (SASA) and secondary structure (SS); and three amino acid features were studied to investigate how amino acid identity can influence protein stability changes: volume, molar mass and hydrophobicity. Results show that SASA values influence directly on stability, since SASA values can suggest if the mutation site is near the hydrophobic core. Secondary structure, however, does not play a significant role on separating neutral mutations from favorable ones according to the results. Molar mass and volume are found to be equivalent and important to stability, however, results on hydrophobicity suggest a low relevance of this feature on protein stability changes. The results are elucidative but further analysis of these descriptors, as well as investigation of new ones, are still necessary to accomplish mutation features prediction.

Palavras-chave: Stability prediction. Site-directed mutagenesis. Statistical analysis.

LISTA DE ILUSTRAÇÕES

Figura 1 – Aminoácidos comuns em proteínas	16
Figura 2 – Variação da energia livre durante o enovelamento	17
Figura 3 – Representação do efeito hidrofóbico	18
Figura 4 – Forças de van der Waals	19
Figura 5 – Ligações de hidrogênio na estabilização de estruturas secundárias	20
Figura 6 – Representação da camada acessível ao solvente	22
Figura 7 – Método de evolução direta	25
Figura 8 – Razão de mutações neutras em intervalos de SASA	36
Figura 9 – Razão de mutações neutras em intervalos de SASA	37
Figura 10 – Categorização quanto à acessibilidade ao solvente	39
Figura 11 – Categorias quanto à acessibilidade ao solvente em cristal de uma proteína globular	40
Figura 12 – Razão de mutações favoráveis em intervalos de SASA	41
Figura 13 – Correlação entre valores de $\Delta\Delta G$ preditos pelo I-Mutant2.0	42
Figura 14 – Razão de mutações favoráveis em intervalos de SASA	43
Figura 15 – Frequência de mutações favoráveis por categoria de SASA	45
Figura 16 – Correlação entre volume e massa molecular	51
Figura 17 – Distribuição de Δ volume por tipo de impacto na estabilidade	52
Figura 18 – Distribuição de Δ volume por tipo de impacto na estabilidade	53
Figura 19 – Δ volume de mutações enterradas	53
Figura 20 – Δ volume de mutações enterradas	54
Figura 21 – Distribuição de Δ hidrofobicidade por tipo de impacto na estabilidade	54
Figura 22 – Δ Hidrofobicidade de mutações enterradas	55

LISTA DE TABELAS

Tabela 1 – Aceptores e doadores de ligações de hidrogênio	20
Tabela 2 – Categorias de mutações quanto à estabilidade	30
Tabela 3 – Descritores analisados	30
Tabela 4 – Categorização adotada quanto à estrutura secundária	32
Tabela 5 – Caracterização dos resíduos de aminoácidos quanto aos parâmetros estudados	33
Tabela 6 – Proporção de mutações neutras em intervalos de SASA	37
Tabela 7 – Intervalos de confiança de SASA	38
Tabela 8 – Qui-quadrado observado para mutações neutras	39
Tabela 9 – Proporção de mutações favoráveis em intervalos de SASA	44
Tabela 10 – Qui-quadrado observado para mutações favoráveis	45
Tabela 11 – Proporção de mutações neutras por SASA e elemento de ES	47
Tabela 12 – Proporção de mutações favoráveis por SASA e elemento de ES	48
Tabela 13 – Mutações favoráveis e desfavoráveis com SASA entre 150 e 250	49
Tabela 14 – χ^2 tabelado	63

LISTA DE ABREVIATURAS E SIGLAS

CH ₃	metil
DNA	ácido desoxirribonucleico
DSSP	<i>define secondary structure of proteins</i>
epPCR	<i>error prone PCR</i>
ES	estrutura secundária
PCR	reação em cadeia da polimerase
PDB	<i>Protein Data Bank</i>
SASA	área superficial acessível ao solvente

LISTA DE SÍMBOLOS

α	Alfa
β	Beta
K	Constante de equilíbrio
R	Constante universal dos gases
G	Energia livre de Gibbs
H	Entalpia
S	Entropia
\ln	Logaritmo neperiano
χ^2	Qui-quadrado
T	Temperatura absoluta
Δ	Varição

SUMÁRIO

1	INTRODUÇÃO	15
1.1	Enovelamento de proteínas	16
1.2	Forças envolvidas no enovelamento	17
1.3	Acessibilidade ao solvente	21
1.4	Estabilidade de proteínas	23
1.5	Mutação <i>in vitro</i> e <i>in silico</i>	24
2	OBJETIVOS	28
2.1	Objetivo Geral	28
2.2	Objetivos Específicos	28
3	METODOLOGIA	29
3.1	O Banco de Dados	29
3.2	Variação da Energia Livre	29
3.3	Descritores Estruturais	30
3.3.1	SASA	31
3.3.2	<i>Estrutura secundária</i>	31
3.4	Descritores de Identidade	32
3.5	Teste Estatístico	32
3.6	Ampliação do Banco de Dados	34
4	RESULTADOS E DISCUSSÃO	35
4.1	Classificação da Mudança de Estabilidade	35
4.2	Descritores Estruturais	36
4.2.1	<i>Acessibilidade ao Solvente</i>	36
4.2.1.1	<i>Mutações Neutras</i>	36
4.2.1.2	<i>Ampliação do Banco de Dados</i>	40
4.2.1.3	<i>Mutações Favoráveis</i>	43
4.2.2	<i>Estrutura secundária</i>	45
4.3	Descritores de identidade	47
4.3.1	<i>Volume e Massa molecular</i>	50
4.3.2	<i>Hidrofobicidade</i>	52
5	CONCLUSÕES E PERSPECTIVAS	56

REFERÊNCIAS	57
ANEXOS	62
ANEXO A – Tabela de Qui-Quadrado	63

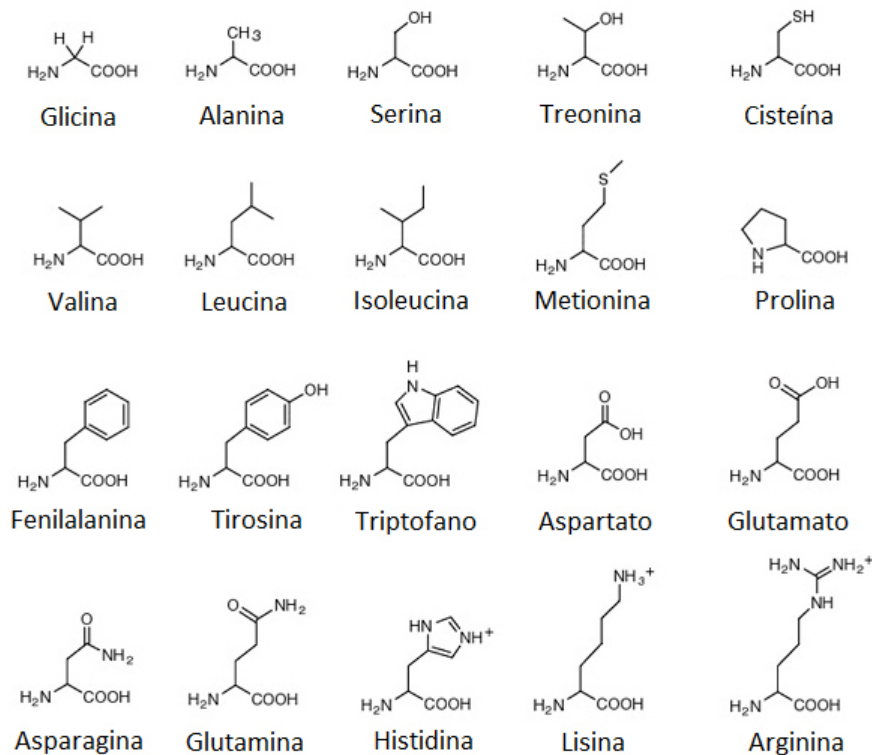
1 INTRODUÇÃO

Proteínas são moléculas de grande interesse comercial pelo seu número crescente de aplicações em praticamente qualquer área de atuação da biotecnologia. O grande potencial de uso dessas moléculas advém de suas numerosas funções biológicas, como suporte estrutural, mobilidade, proteção, regulação, catálise ou transporte (MÁRQUEZ-CHAMORRO *et al.*, 2015). O mercado biofarmacêutico é o mais expressivo dentre as aplicações de proteínas e é dominado por vacinas, enzimas e anticorpos monoclonais, sendo atualmente avaliado em US\$70 bilhões, apenas 30 anos após o primeiro anticorpo monoclonal ser comercializado (BUYEL *et al.*, 2017). As proteínas mais comumente utilizadas pelas indústrias são aquelas que promovem atividade catalítica (enzimas), que possuem aplicação em diversos setores, como na produção de detergentes, alimentos, ração animal, tecidos, couro, produtos químicos, polpa e papel, sendo os mercados de detergentes e alimentos os de maior relevância econômica (DEMAIN, 2007). O uso de enzimas resulta em benefícios para o processo e para o meio ambiente, uma vez que são biodegradáveis e substituem compostos sintéticos que trazem prejuízos ecológicos (WANG, 1999), além de contribuírem para a redução da emissão de subprodutos poluentes (STAHLKE, 2015).

Em 2015, o mercado global de enzimas industriais era avaliado em 4 bilhões de dólares, sendo estimada em 240 milhões de dólares a contribuição do mercado brasileiro (6% do mercado global) (STAHLKE, 2015). O mercado de proteínas tende a crescer e a contínua melhoria das propriedades dessas moléculas é fundamental para superar limitações, como baixa estabilidade, que impossibilita o uso de proteínas em condições de pH e temperatura extremas, comuns em processos industriais (BALCÃO; VILA, 2015).

A estabilização proteica pode ser atingida por imobilização da proteína em um suporte, adição de um ligante estabilizador, tratamento com substâncias que modificam o sistema ou por mutação (OLIVEIRA; MANTOVANI, 2009). Quando uma mutação ocorre na região de expressão de uma proteína, esta pode sofrer alterações em suas propriedades, como capacidade de interação com outras moléculas, localização subcelular, expressão, função na célula, enovelamento e estabilidade, sendo estes dois últimos analisados posteriormente neste trabalho (REVA *et al.*, 2011). Essas alterações ocorrem por conta de mudanças na sequência de resíduos de aminoácidos que compõem a proteína, por conta da variedade de características que os aminoácidos possuem, decorrente de sua cadeia lateral única (Figura 1).

Figura 1 – Aminoácidos comuns em proteínas



Dentre os α -aminoácidos que existem, apenas 20 são comumente encontrados na constituição de proteínas e se diferenciam por suas cadeias laterais, ligadas ao carbono- α , que atribuem a essas moléculas propriedades bioquímicas e funções diferentes. A diversidade de características dos aminoácidos é o que possibilita a existência de proteínas com funções tão distintas.

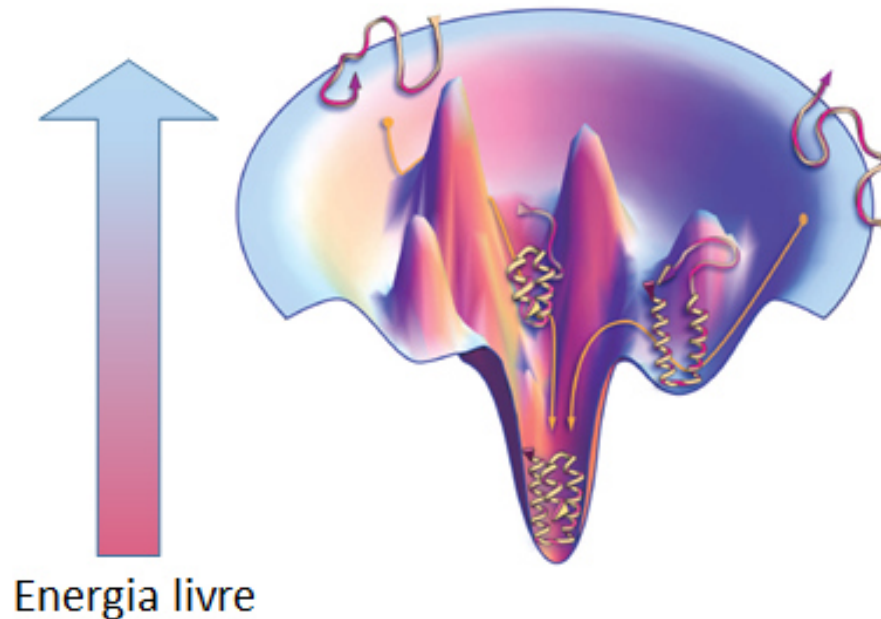
Fonte – Sturm (2015).

1.1 Enovelamento de proteínas

Proteínas recém-sintetizadas sofrem um processo de modificação conformacional, chamado de enovelamento, até atingirem a conformação tridimensional estável em que se encontra no seu ambiente de atuação, chamada de conformação nativa (PACE *et al.*, 2009). O estado enovelado compreende um número limitado de conformações preferencialmente adotadas pela proteína, sendo a conformação nativa geralmente o estado de menor energia (WANG, 1999).

O enovelamento de proteínas ocorre naturalmente após sua síntese ou com o auxílio de outras proteínas, as chamadas chaperonas. Independentemente do modo pelo qual a proteína se enovela, modificações dos ângulos entre ligações na cadeia principal e nas cadeias laterais ocorrem como consequência de interações entre os átomos presentes no sistema proteína-solvente (DILL, 1990), de modo que a sequência primária de uma proteína contém toda a informação necessária para que o enovelamento ocorra espontaneamente (Figura 2) (KUMAR, 2009).

Figura 2 – Variação da energia livre durante o enovelamento



A espontaneidade de qualquer processo é determinada na termodinâmica em termos de energia livre ($\Delta G = \Delta H - T \Delta S$, $\Delta G < 0$ indica que o processo pode ocorrer de forma espontânea). Onde ΔH é variação de entalpia, T é a temperatura absoluta e ΔS é variação de entropia. Entalpia se refere à energia liberada (ou consumida) em forma de calor, que a nível molecular está relacionada à formação ($\Delta H > 0$) ou quebra ($\Delta H < 0$) de ligações. Entropia é uma grandeza que mede a tendência natural de um sistema a se desordenar ($\Delta S > 0$ sempre), resultando em aumento do grau de liberdade (número de microestados possíveis para o macroestado em que o sistema se encontra) (OCHS, 1996).

Fonte – adaptado de Dill e MacCallum (2012).

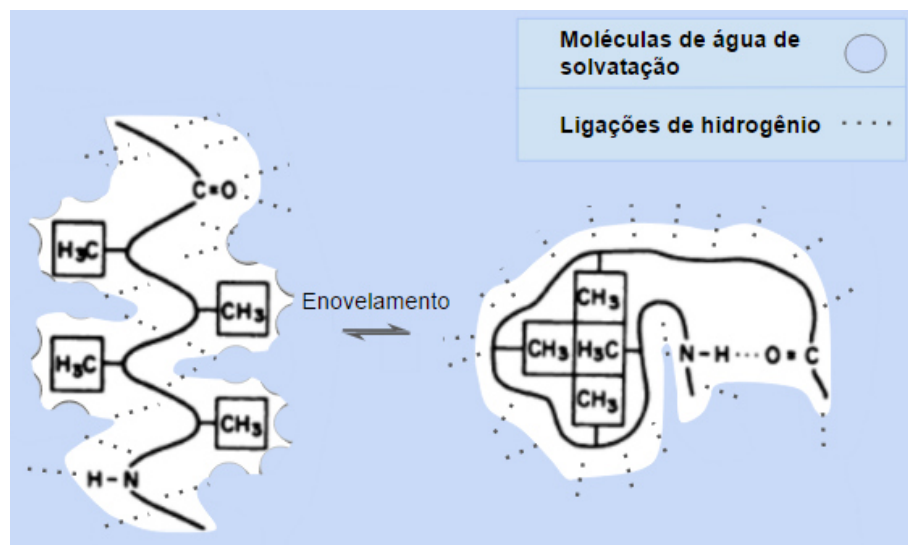
As interações envolvidas no enovelamento incluem propensões termodinâmicas (efeito hidrofóbico), formação de pontes salinas, repulsão eletrostática, ligações dissulfeto, interações com componentes do solvente, ligações de hidrogênio e outras interações de natureza eletrostática (WANG, 1999). Tais forças são resultados de interações entre átomos pertencentes à proteína e entre átomos da proteína e do solvente.

1.2 Forças envolvidas no enovelamento

A contribuição mais expressiva que impulsiona o enovelamento de proteínas é geralmente atribuída ao efeito hidrofóbico, que resulta na formação de um centro apolar parcialmente ou completamente desidratado (WANG, 1999). O efeito hidrofóbico é um fenômeno entrópico, ocorrendo naturalmente por diminuir o nível de organização do sistema – ou aumentar o grau de liberdade de parte do sistema (CHANDLER, 2005). Moléculas de água se organizam em uma camada de hidratação em contato com grupos hidrofóbicos, chamada camada solvatação, quando

esses grupos são afastados do solvente o grau de liberdade das moléculas de água aumenta, diminuindo a ordem do sistema (BIEDERMANN *et al.*, 2014) (Figura 3). O efeito hidrofóbico é facilmente identificado em fenômenos macroscópicos e microscópicos comuns como a separação espontânea de óleo e água e na formação de micelas a partir moléculas anfipáticas em solução (GREAVES; DRUMMOND, 2013).

Figura 3 – Representação do efeito hidrofóbico



Radicais metil (CH_3) representam domínios hidrofóbicos que se agrupam e evitam a formação de água de solvatação.

Fonte – adaptado de Pace *et al.* (1996).

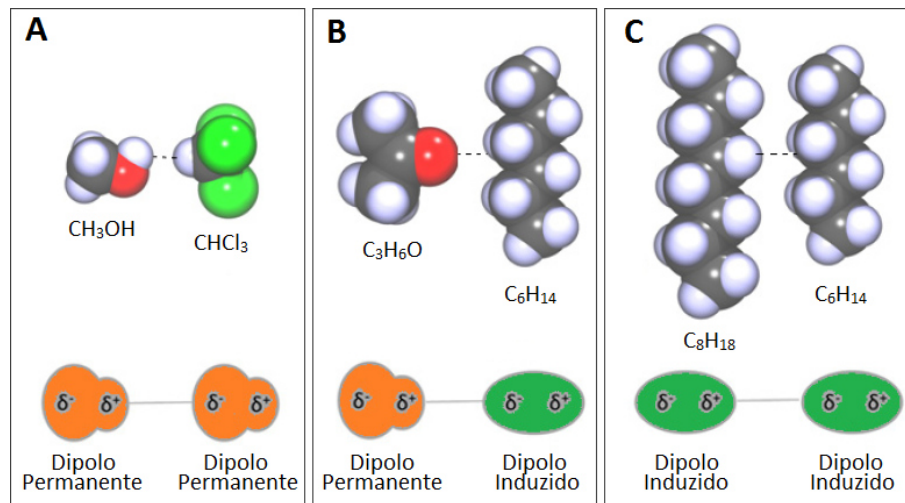
Em proteínas, o efeito hidrofóbico é também atribuído a um fator entálpico. Durante o enovelamento, o centro apolar tende a desidratar pois a água que se encontra entre grupos apolares pode formar mais ligações de hidrogênio ao serem liberadas para se juntar ao solvente do sistema (BIEDERMANN *et al.*, 2014). A formação das ligações de hidrogênio gera aumento de entalpia e favorece a espontaneidade do enovelamento (CHANDLER, 2005).

Em geral, durante o enovelamento, cerca de 80% dos resíduos de cadeia lateral apolar (alanina, valina, isoleucina, leucina, metionina, fenilalanina, triptofano e cisteína) se agrupam ao centro da proteína, fora do contato com a água (WANG, 1999). Resíduos polares podem estar presentes no centro hidrofóbico da conformação nativa, porém devem estar envolvidos em interações eletrostáticas favoráveis (como ligações de hidrogênio ou pontes salinas) (HENDSCH; TIDOR, 1994).

Forças de van der Waals são interações relacionadas à distribuição e movimentação da nuvem eletrônica de moléculas polares e apolares resultando em interações atrativas e repulsivas

(BUTT *et al.*, 2005). Forças de van der Waals podem surgir a partir da interação envolvendo dipolos fixos ou dipolos induzidos. São definidas como o somatório de forças de curto alcance de Keesom (dipolo-dipolo), Debye (dipolo-dipolo induzido) e London (dipolo instantâneo-dipolo instantâneo) (Figura 4). Forças de van der Waals possuem grande relevância para o enovelamento de proteínas por serem interações que ocorrem entre quaisquer duas moléculas (BERLAND *et al.*, 2015).

Figura 4 – Forças de van der Waals



Em A, força de Keesom. Em B, força de Debye. Em C, força de London. Forças envolvendo dipolos permanentes são normalmente de maior intensidade.

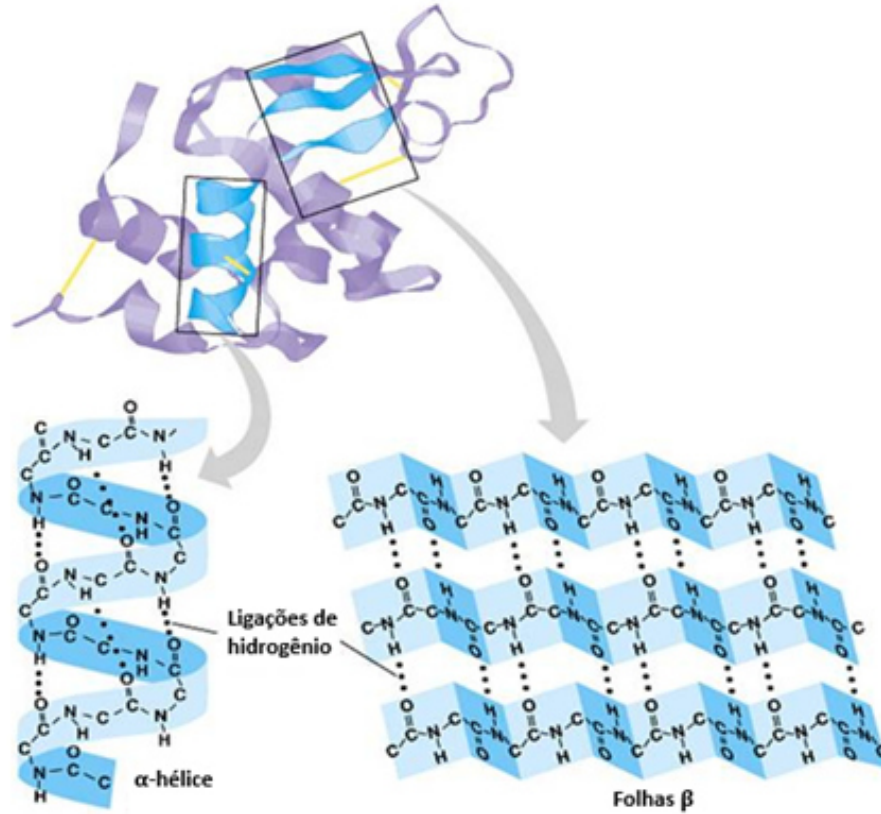
Fonte – adaptado de Leite *et al.* (2012).

Ligações de hidrogênio são fortes interações dipolo-dipolo entre um hidrogênio ligado covalentemente a um átomo (chamado de doador) e outro átomo de grande eletronegatividade, como flúor, oxigênio ou nitrogênio (chamado de receptor). A força de uma ligação de hidrogênio depende da eletronegatividade e da orientação do doador e receptor da ligação. Ligações de hidrogênio são mais fortes que outras interações dipolo-dipolo porque átomos de hidrogênio têm forte tendência de se tornarem polarizados positivamente e possuem um raio atômico muito pequeno (BRYANT; MCCLEMENTS, 1998).

Todos os resíduos de aminoácido possuem grupos doadores e aceptores de ligações de hidrogênio nos átomos da cadeia principal, e alguns os possuem também na cadeia lateral, podendo formar ligações entre si ou com moléculas de água do solvente. Ligações de hidrogênio formadas entre grupos amina e grupos carboxila da cadeia principal são interações importantes para a estabilização de estruturas secundárias (Figura 5) (BAKER; HUBBARD, 1984). Os grupos funcionais presentes na cadeia lateral dos aminoácidos, que permite que formem ligações

de hidrogênio são indicados na Tabela 1.


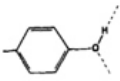
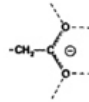
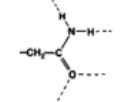
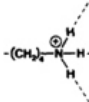
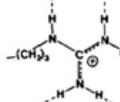
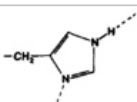
Figura 5 – Ligações de hidrogênio na estabilização de estruturas secundárias



As ligações de hidrogênio que estabilizam estruturas secundárias não envolvem átomos das cadeias laterais.

Fonte – adaptado de Reece *et al.* (2015).

Tabela 1 – Aceptores e doadores de ligações de hidrogênio

Grupos funcionais	Resíduos	Grupos funcionais	Resíduos
Hidroxila	 Serina Treonina	Fenol	 Tirosina
Carboxila	 Aspartato Glutamato	Carboxiamida	 Asparagina Glutamina
Amina	 Lisina	Guanidina	 Arginina
Imidazol	 Histidina		

Fonte – adaptado de Reece *et al.* (2015).

Ligações dissulfeto são ligações covalentes que se formam através da reação de oxidação entre cadeias laterais de dois resíduos de cisteína, conectando fortemente resíduos que são distantes na estrutura primária. Apesar de serem ligações fortes, não contribuem intensamente para a estabilização da proteína, pois não são muito frequentes, sendo formadas apenas após os resíduos de cisteína serem aproximados por conta de outras forças que regem o enovelamento (BRYANT; MCCLEMENTS, 1998).

A presença de partículas carregadas eletronicamente na estrutura de proteínas influencia na estabilidade molecular devido a efeitos eletrostáticos específicos e inespecíficos. O efeito inespecífico é a repulsão existente quando há acúmulo de cargas na proteína. Uma grande densidade de cargas leva a forças repulsivas mais intensas e maior pressão energética de desnaturação, pois o estado desnaturado se torna mais estável por ocupar maior volume, diminuindo a densidade de cargas (DILL, 1990). Efeitos eletrostáticos específicos são regidos pela lei de Coulomb e ocorrem quando cadeias laterais com cargas iguais ou opostas estão próximas, formando interações eletrostáticas repulsivas ou atrativas (pontes salinas), respectivamente (BRYANT; MCCLEMENTS, 1998).

Apesar do enovelamento de proteínas ser um processo espontâneo ($\Delta G_{\text{resultante}} < 0$), existem também propulsões termodinâmicas às quais proteínas estão sujeitas que são desfavoráveis ao enovelamento. Tais propulsões estão associadas ao custo entrópico ocasionado pela diminuição do grau de liberdade que o polímero sofre durante as etapas do enovelamento. Como o enovelamento resulta em um agregado molecular, pouco espaço se torna disponível para as cadeias laterais e as inúmeras interações que são criadas diminuem a liberdade da cadeia principal (DILL, 1990).

1.3 Acessibilidade ao solvente

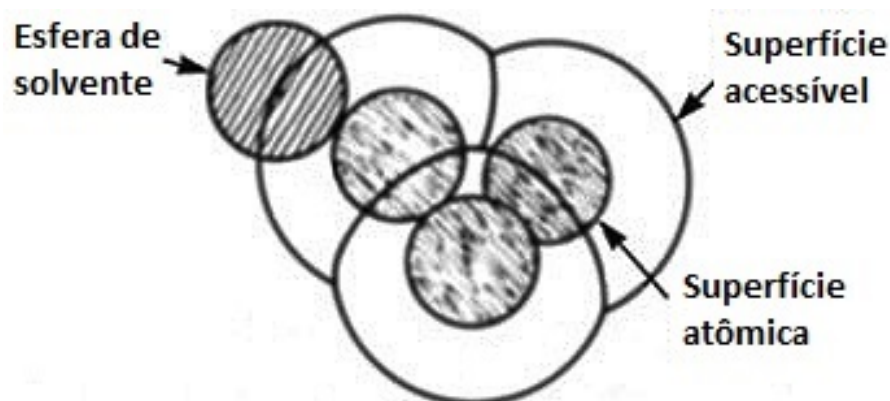
Durante o enovelamento, para obter-se uma estrutura estável em um ambiente aquoso, domínios apolares devem ser preferencialmente mantidos agrupados no interior, afastados da superfície, enquanto que domínios polares devem ser mantidos expostos ao solvente, como é observado em micelas de ácidos graxos que se formam naturalmente em meio polar. Em relação a conformações nativas de proteínas, as áreas polar e apolar não são tão facilmente distinguíveis como em micelas, pois, após o enovelamento, podem ser identificadas regiões de maior ou menor contato com o solvente, derivando-se, assim, uma grandeza contínua para quantificar o nível de contato de resíduos em cada posição com moléculas do solvente.

A área superficial acessível ao solvente (SASA) é um parâmetro que estima quantitativamente o nível de exposição de um aminoácido às moléculas do meio em que uma proteína se encontra. O conhecimento da SASA de resíduos de uma proteína é importante para a determinação de sua estrutura tridimensional pois, durante o enovelamento de proteínas, a compactação máxima da estrutura ocorre como resultado da aversão de alguns resíduos ao solvente (SHAHEEN *et al.*, 2015).

A determinação de quais resíduos se encontram enterrados ou expostos auxilia em pesquisas como a predição do ambiente subcelular de uma proteína através da análise dos resíduos expostos (ANDRADE *et al.*, 1998), predição de sítios de interação com outras moléculas através da identificação de regiões expostas com baixa polaridade, úteis na formação de interfaces hidrofóbicas (CHEN; ZHOU, 2005a), e elucidação de características estruturais de proteínas – um exemplo é a grande contribuição do núcleo hidrofóbico para a estabilidade, reforçada pelo fato de a maioria das mutações deletérias ocorrerem em regiões enterradas de proteínas (CHEN; ZHOU, 2005b). A SASA é, também, uma atribuição comum em modelos de predição de estabilidade e rotinas de dinâmica molecular.

A SASA de uma molécula é definida como a superfície acessível da esfera concêntrica à molécula, cujo raio é a soma do raio da molécula e o raio de uma esfera de solvente (RICHMOND, 1984) (Figura 6). Uma representação esférica de solvente de raio 1,4 Å é utilizada como sonda ao redor da superfície atômica e a área de superfície acessível é contabilizada se a sonda não colidir com outros átomos da vizinhança (KABSCH; SANDER, 1983).

Figura 6 – Representação da camada acessível ao solvente



Fonte – Richmond (1984)

1.4 Estabilidade de proteínas

O enovelamento de proteínas é um processo reversível que pode ser modulado pelas condições ambientais da molécula, dando à estabilidade de proteínas um conceito abstrato, dependente das condições em que se encontra a proteína. Dessa forma, uma definição de estabilidade, de forma prática, representa a capacidade de uma proteína de estabelecer sua função, dado o conjunto de condições físicas e químicas a que é submetido (DAY, 1996). Para a termodinâmica, estabilidade é entendida como a variação de energia livre (ΔG) resultante do processo de enovelamento.

O processo de enovelamento tem início ainda durante a síntese de uma proteína, o que dificulta a determinação experimental do ΔG de enovelamento, pela grande quantidade de interferentes (HARTL; HAYER-HARTL, 2009). Assim, partindo do princípio de que o processo de enovelamento é antagônico ao processo de desnaturação, os métodos experimentais para quantificar o ΔG consistem no monitoramento da desnaturação induzida de uma proteína, por desnaturação química ou térmica (LEE; BLABER, 2009). O monitoramento durante a desnaturação é feito através de técnicas de calorimetria ou espectroscopia, de dicroísmo circular ou por fluorescência, por exemplo, e os dados gerados são obtidos em termos da constante de equilíbrio da reação de enovelamento ($K_{enovelamento}$), que possibilita a determinação de ΔG através da Equação 1.1 (HUYGHUES-DESPOINTES *et al.*, 2001).

$$\Delta G = -RT \ln K_{enovelamento}, \quad (1.1)$$

onde R é a constante universal dos gases ($R = 1,987 \text{ cal/mol K}$) e T é a temperatura absoluta (PACE *et al.*, 2009).

O ΔG de enovelamento de uma proteína globular é relativamente pequeno, na ordem de 5 kcal/mol a 15 kcal/mol, quando comparado com, por exemplo, a energia de formação de uma ligação covalente (aproximadamente 30 kcal/mol a 100 kcal/mol) (DAY, 1996).

A estabilidade proteica é um reflexo de características estruturais específicas. Proteínas termofílicas, por exemplo, são mais estáveis e possuem uma estrutura rígida, devido a um maior número de interações não-covalentes (pontes salinas, ligações de hidrogênio e forças de van der Waals) e covalentes (pontes dissulfeto), empacotamento hidrofóbico compacto, menor presença de grupos hidrofóbicos na superfície e maior quantidade de prolina (SIDDIQUI, 2015). Tais características causam aumento da estabilidade por diminuir a entropia da molécula.

Uma maneira de obter proteínas com estabilidade melhorada é através de proteínas

homólogas isoladas de extremófilos (termófilos, hipertermófilos, halófilos, acidófilos, alcalófilos, piezófilos metalotolerantes e psicrofilos), porém o isolamento desse tipo de microrganismo não é um processo simples (DEY *et al.*, 2016) e poucos possuem naturalmente uma atividade enzimática satisfatória para a indústria ou se mantêm estáveis em um grande intervalo de temperatura (SIDDIQUI, 2015). A produção de proteínas mutantes, por sua vez, é uma tecnologia que evolui cada vez mais e se torna cada vez mais exata.

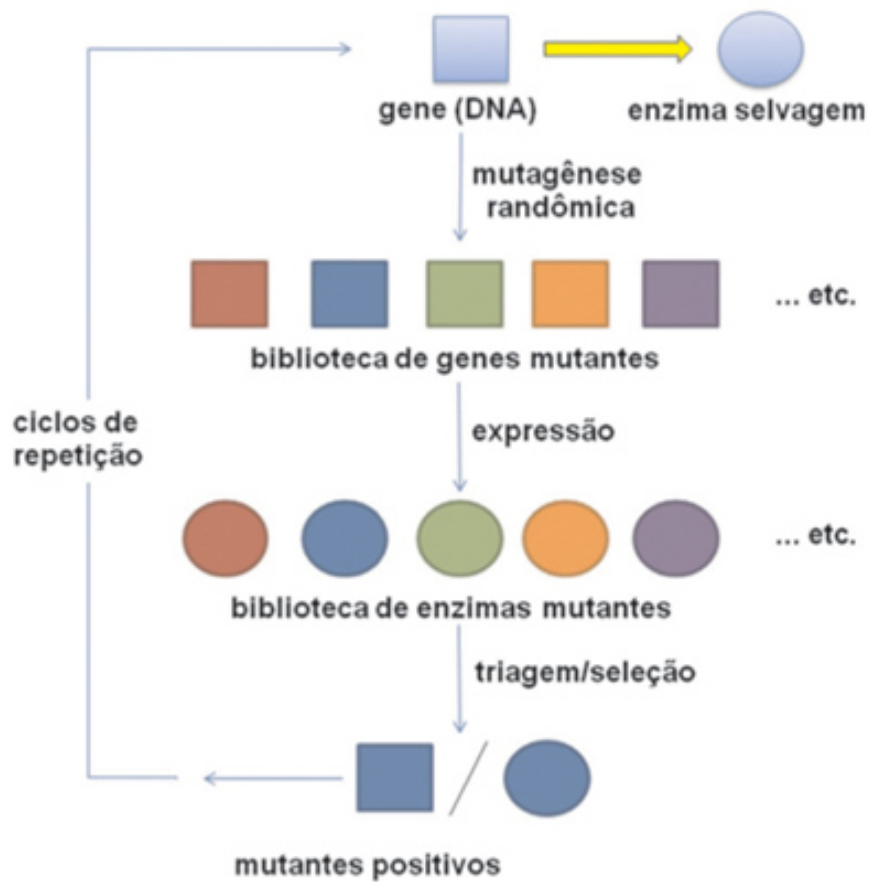
1.5 Mutação *in vitro* e *in silico*

As técnicas utilizadas para obtenção de mutantes *in vitro* se resumem à indução de mutações aleatórias ou a mutação sítio-dirigida. A indução de mutações aleatórias, ou evolução direta, é um método baseado no processo natural de evolução, tendo como objetivo gerar diversidade genética e selecionar mutantes com características desejadas (Figura 7) (OLIVEIRA; MANTOVANI, 2009).

A mutagênese aleatória é produzida por modificações causadas ao ácido desoxirribonucleico (DNA) durante sua replicação *in vitro* por reação em cadeia da polimerase (PCR). Mutagênese aleatória é conduzida principalmente através de *error prone PCR* (epPCR), embaralhamento de DNA e mutagênese de saturação (REETZ, 2017). A epPCR se baseia na propriedade de algumas polimerases termoestáveis em apresentarem baixa fidelidade ao longo da extensão da cadeia de DNA introduzindo erros ao acaso (OLIVEIRA; MANTOVANI, 2009). O embaralhamento de DNA (ou DNA *shuffling*) utiliza de oligonucleotídeos gerados pela hidrólise aleatória de um grupo de genes homólogos. Esses oligonucleotídeos participam de ciclos de PCR sem *primers* onde o produto é gerado pelo anelamento aleatório desses oligopeptídeos e posterior extensão das cadeias híbridas (SILVA *et al.*, 2003). A mutagênese por saturação é feita utilizando *primers* degenerados e não é um método tão aleatório quanto os outros, pois a posição da mutação é determinada previamente. Assim, a aleatoriedade do método decorre apenas em relação à identidade do aminoácido a ser inserido no local da mutação, gerando os 19 mutantes possíveis (REETZ, 2011).

Em mutagênese sítio-dirigida, mutações pontuais são inseridas através de modificações controladas na sequência do gene de interesse (SANTOS, 2005). O método de mutação sítio-dirigida é similar à mutagênese de saturação, com a diferença de que a identidade do aminoácido é fixa e predeterminada também, ou seja, os *primers* utilizados não são degenerados e trazem consigo os nucleotídeos selecionados para dar origem à proteína mutante (MITCHELL

Figura 7 – Método de evolução direta



A evolução direta consiste em mutagênese randômica do gene de interesse, criação de bibliotecas de genes e produtos mutantes e, por fim, a seleção dos mutantes adaptados. Repetições do processo garantem maior variabilidade genética.

Fonte – adaptado de Oliveira e Mantovani (2009).

et al., 2012. No prelo).

A criação de bibliotecas de mutantes geradas por mutagênese aleatória resulta em uma extensa quantidade de informação e o procedimento depende da análise de até milhares de colônias mutantes, resultando num processo caro e demorado. Após a identificação de 6 posições de interesse, por exemplo, existem ainda por volta de 64 milhões (20^6-1) de mutantes que podem ser criados (DEUTSCH; KRISHNAMOORTHY, 2007). Estudos reportaram que as mutações mais comuns em bibliotecas de mutagênese aleatória são deletérias, com uma frequência superior a um terço de todos os mutantes produzidos, enquanto mutações benéficas ocorrem em frequências por volta de apenas 0,1% (ITO *et al.*, 2013).

Assim, mutagênese sítio-dirigida é preferencialmente utilizada para redução de custos, principalmente se aliada a tecnologia computacional (ITO *et al.*, 2013). Métodos de bioinformática utilizam uma variedade de informações, resultantes de experimentos preliminares,

modelagem computacional ou estudo de sequências homólogas para reduzir o número de aminoácidos e posições investigados com os quais a mutação deve ser preferencialmente conduzida (BOMMARIUS *et al.*, 2006).

Mutações sítio-dirigidas podem ser simuladas *in silico* através de programas que substituem um aminoácido, de posição especificada, por outro aminoácido, respeitando o volume ocupado por cada átomo da estrutura para evitar superposição dos átomos. Informações estruturais da conformação nativa de proteínas podem ser acessadas através do Protein Data Bank (PDB), um banco de dados com estruturas proteicas descritas em detalhamento atômico (BERMAN *et al.*, 2000).

Como todo aminoácido desempenha múltiplas funções em proteínas, a substituição de um resíduo por outro, para melhoria de uma característica, provavelmente afetará negativamente a proteína em outro aspecto (STUDER *et al.*, 2014). Assim, para melhorar a estabilidade de uma enzima que participa de um processo industrial, a fim de que resista a maiores temperaturas, por exemplo, cuidados devem ser tomados para que outros fatores não sejam prejudicados, como a redução de sua afinidade pelo substrato. Da mesma forma, se uma mutação em um peptídeo utilizado para fins terapêuticos tem como objetivo alterar uma sequência facilmente reconhecida por células B do sistema imunológico humano, é preciso garantir que, além da diminuição da alergenicidade, o mutante também permaneça com sua estabilidade inalterada.

Quanto à estabilidade, dois tipos de resultados são desejados após a mutagênese: produção de mutante mais estável que a proteína selvagem, ou seja, com atividade conservada em ambientes mais extremos; ou mutações que resultem em impacto neutro na estabilidade (ou apenas não desestabilizador), garantindo que o mutante seja tão estável quanto a proteína selvagem, em trabalhos que visam a melhoria do mutante em outro aspecto.

Uma vez estabelecido o objetivo da mutação, é necessário conhecer a posição em que deve ocorrer e por qual aminoácido o resíduo original deve ser substituído. Porém, fazer tais determinações com precisão é uma tarefa complexa por conta da pluralidade de funções exercidas por cada aminoácido, de modo que a melhor combinação posição-aminoácido inserido deve ser selecionada dentro de um conjunto de opções que é muitas vezes extenso e o resultado de cada combinação, apenas observável *in vitro*.

Assim, uma maneira de prever o conjunto mínimo de mutações que devem ser preferencialmente testadas *in vitro* para melhor atender ao objetivo escolhido é de grande importância para a redução de tempo e gastos com etapas de experimentação.

Este estudo teórico-estatístico foi realizado com o objetivo de prever características associadas ao conjunto de mutações que possua a maior chance de alcançar o objetivo esperado para a mutagênese, reduzindo o espaço de busca e gastos na etapa de experimentação *in vitro*.

2 OBJETIVOS

2.1 Objetivo Geral

Entender o envolvimento de parâmetros estruturais e da identidade de aminoácidos em determinar mudanças na estabilidade conformacional de proteínas após mutações pontuais de substituição.

2.2 Objetivos Específicos

- Classificação de mutações com base na mudança de estabilidade.
- Classificação da acessibilidade ao solvente em camadas e de elementos de estrutura secundária.
- Análise das proporções de mutantes neutros e favoráveis para cada descritor e para combinações de descritores.
- Avaliação da relevância dos descritores para a predição de mudanças na estabilidade após mutação.

3 METODOLOGIA

3.1 O Banco de Dados

O banco de dados utilizado para as análises estatísticas compreende três conjuntos obtidos experimentalmente e curados manualmente que contêm 1005 (GUEROIS *et al.*, 2002), 2154 (POTAPOV *et al.*, 2009) e 1210 (KELOGG *et al.*, 2011) mutações pontuais e suas respectivas variações de energia livre (CALIFORNIA, 2017). Cada entrada do banco inclui o código PDB (BERMAN *et al.*, 2000) da proteína selvagem, posição da mutação e resíduo inserido, estrutura secundária e $\Delta\Delta G$. Às mutações que possuíam mais que um valor de $\Delta\Delta G$ associado foi atribuída a média dos dados. Todas as mutações presentes no banco de dados correspondem a uma mutação de substituição, portanto, mutações de adição ou deleção de aminoácidos não são contempladas.

3.2 Variação da Energia Livre

A variação da variação de energia livre ($\Delta\Delta G$) contida no banco de dados foi utilizada para determinar o impacto das mutações na estabilidade das estruturas. O $\Delta\Delta G$ é a diferença entre as variações de energia livre da estrutura mutante e da estrutura selvagem (Equação 3.1). A medição experimental de $\Delta\Delta G$ é feita através da determinação do ΔG de enovelamento, antes e após a mutação.

$$\Delta\Delta G_{mutacao} = \Delta G_{mutante} - \Delta G_{selvagem} \quad (3.1)$$

Uma mutação pode ter diversos efeitos na estabilidade da estrutura de uma proteína. Mutações que não impactem negativamente a estrutura da proteína são consideradas *neutras*, mutações que comprometem a estabilidade de uma proteína são consideradas *desfavoráveis* e mutações que aumentam a estabilidade de uma proteína ou tornam o processo de enovelamento espontâneo são chamadas *favoráveis*. Diante da variação contínua de $\Delta\Delta G$ para as mutações presentes no banco de dados, valores de corte devem ser escolhidos para que uma determinada mutação seja considerada como parte de uma das três categorias encontradas na literatura (Tabela 2). Três formas de classificação são propostas inicialmente. A primeira permite variação de até $\pm 0,5$ kcal/mol para uma mutação ser considerada neutra, sendo qualquer variação acima de 0,5 kcal/mol considerada desfavorável e abaixo de -0,5 kcal/mol favorável. A segunda é semelhante à primeira, mas adota o valor máximo de variação de $\pm 1,0$ kcal/mol para uma

mutação ser considerada neutra. Finalmente, a última classificação não contém a categoria de mutação neutra, e divide as mutações apenas em favoráveis e desfavoráveis.

Tabela 2 – Categorias de mutações quanto à estabilidade

Classe	Dado experimental (kcal/mol)	Número de mutantes por classe	Referências
Favorável	$\Delta\Delta G \leq 0$	830 (23%)	Potapov <i>et al.</i> (2009)
Desfavorável	$\Delta\Delta G > 0$	2775 (77%)	Capriotti <i>et al.</i> (2005)
Favorável	$\Delta\Delta G \leq -0,5$	441 (12%)	Capriotti <i>et al.</i> (2008)
Neutro	$-0,5 < \Delta\Delta G < 0,5$	995 (28%)	
Desfavorável	$\Delta\Delta G \geq 0,5$	2169 (60%)	
Favorável	$\Delta\Delta G \leq -1$	243 (7%)	Khan e Vihinen (2010)
Neutro	$-1 < \Delta\Delta G < 1$	1672 (46%)	
Desfavorável	$\Delta\Delta G \geq 1$	1690 (47%)	

Fonte – elaborada pelo autor.

3.3 Descritores Estruturais

Foram investigados 5 descritores em relação à mudança de estabilidade após mutação: dois descritores estruturais (estrutura secundária e acessibilidade ao solvente) e quatro descritores dependentes da identidade dos aminoácidos envolvidos nas substituições (volume, massa molecular e hidrofobicidade).

Tabela 3 – Descritores analisados

Descritor	Tipo
Acessibilidade ao solvente (SASA)	Estrutura
Estrutura secundária	Estrutura
Volume	Identidade
Massa molar	Identidade
Hidrofobicidade	Identidade

Fonte – elaborada pelo autor.

As informações estruturais analisadas nos sítios de mutação foram obtidas através do programa STRIDE (HEINIG; FRISHMAN, 2004). O método do STRIDE se baseia em nos ângulos e distâncias da cadeia principal para classificar as estruturas secundárias, e baseia-se o método geométrico de agrupamento da cadeia lateral em malhas para a determinação SASA (EISENHABER *et al.*, 1995).

3.3.1 SASA

O $\Delta\Delta G$ foi analisado através da frequência de mutações neutras e favoráveis em diferentes intervalos de SASA. Inicialmente, 251 intervalos foram analisados, todos com início em SASA = 0 Å² e com fim variando entre os valores inteiros de 0 a 250. Dessa forma o primeiro intervalo, de 0 a 0, inclui os valores de SASA iguais a 0 Å², o segundo intervalo inclui todas as mutações com SASA maior ou igual a 0 Å² e menor ou igual a 1 Å², o intervalo seguinte variou de 0 Å² a 2 Å², e assim por diante.

Em seguida, foram analisados 9 intervalos diferentes de SASA de comprimento igual a 50 Å² e com início variando a cada 25 Å², de modo que o primeiro intervalo teve início em 0 Å² e fim em 50 Å², o segundo, início em 25 Å² e fim em 75 Å² e assim por diante.

As proporções de mutações neutras e favoráveis foram observadas para cada intervalo de SASA analisados. A representatividade das classes foi validada através de teste em triplicata em que 20% dos dados eram aleatoriamente removidos do banco de dados para integrar um grupo-teste. As frequências de mutações neutras e favoráveis para os 80% dos dados restantes eram analisadas e comparadas às frequências obtidas dentro do conjunto-teste. A validação da hipótese de representatividade das classes foi feita pelo teste do qui-quadrado.

3.3.2 Estrutura secundária

A estrutura secundária é resultado da interação de resíduos através de ligações de hidrogênio. Aminoácidos variam na capacidade de formar elementos de estrutura secundária. As estruturas secundárias dominantes são α -hélices e folhas- β , porém existem outros tipos de hélices, como a hélice 3_{10} e a hélice π . Outras regiões são caracterizadas por não se diferenciarem em um elemento bem definido e são agrupadas num grupo genérico de estruturas irregulares que contém laços, voltas, pontes etc., chamado de *coil* (HOFFMANN *et al.*, 2014).

A classificação de estrutura secundária adotada abrange três grupos contendo elementos distintos (classificados de acordo com o algoritmo *define secondary structure of proteins* (DSSP) (TOUW *et al.*, 2015)) e um quarto grupo que une os elementos de estrutura secundária considerados definidos (Tabela 4). Os mutantes estão distribuídos de forma semelhante entre os elementos de estrutura secundária e possuem dados suficientes que permitam a análise dos grupos.

A avaliação da estrutura secundária foi feita observando-se a frequência de mutações

Tabela 4 – Categorização adotada quanto à estrutura secundária

Classe	Estruturas compreendidas	Número de mutantes por classe
Hélice	α -hélice, hélice 3_{10} e a hélice π	1288 (36%)
Folha- β	Folha- β	1112 (31%)
<i>Coil</i>	Voltas, dobras, ponte- β , estruturas irregulares	1205 (33%)
Definida	Hélice + folha- β	2400 (67%)

Fonte – elaborada pelo autor.

neutras e favoráveis que pertencem a determinada classe de estrutura secundária e, em seguida, analisando cada interseção entre as classes de estrutura secundária e de SASA, de modo que 16 subclasses (4 classes de SASA x 4 classes de estrutura secundária) foram consideradas, como é o exemplo da subclasse de mutações que pertencem a uma α -hélice e têm SASA entre 25 \AA^2 e 50 \AA^2 .

3.4 Descritores de Identidade

A atribuição de cada resíduo quanto aos descritores da identidade do aminoácido foram retirados da literatura (Tabela 5). A relação entre cada descritor e $\Delta\Delta G$ foi calculada a partir da diferença absoluta entre o valor atribuído ao resíduo mutante e o valor atribuído ao resíduo selvagem.

Em uma segunda análise, a classificação dos dados leva em conta apenas mutações com valor de SASA entre 0 \AA^2 e 25 \AA^2 (grupo considerado enterrado no interior da proteína), por ser a região que possui maior densidade de dados e proporção mais equilibrada entre mutações neutras, favoráveis e desfavoráveis (30%, 35% e 35%, respectivamente).

3.5 Teste Estatístico

O qui-quadrado (χ^2) é um teste de hipóteses que se destina a avaliar a associação existente entre variáveis qualitativas. É um teste não paramétrico, ou seja, não depende de parâmetros populacionais, como média e variância. O princípio básico deste método é comparar as possíveis divergências entre as frequências observadas e esperadas para um certo evento e decidir se os dados concordam ou não (PARANÁ, 2011). Para comparar as proporções obtidas para as camadas de SASA, os dois subgrupos formados aleatoriamente foram testados: o subgrupo contendo 80% dos dados deu origem às frequências esperadas e o subgrupo com 20% dos dados originou as frequências observadas.

Tabela 5 – Caracterização dos resíduos de aminoácidos quanto aos parâmetros estudados

Aminoácidos	Massa molar (g/mol)	Volume (Å ³)	Volume (Å ³)	Hidrofobicidade (kcal/mol)
Alanina (A)	71,09	67,0	88,6	0,17
Arginina (R)	156,19	148,0	173,4	0,81
Asparagina (N)	114,11	96,0	114,1	0,42
Aspartato (D)	115,09	91,0	111,1	1,23
Cisteína (C)	103,15	86,0	108,5	-0,24
Glutamina (Q)	128,14	114,0	143,8	0,58
Glutamato (E)	129,12	109,0	138,4	2,02
Glicina (G)	57,05	48,0	60,1	0,01
Histidina (H)	137,14	118,0	153,2	0,96
Isoleucina (I)	113,16	124,0	166,7	-0,31
Leucina (L)	113,16	124,0	166,7	-0,56
Lisina (K)	128,17	135,0	168,6	0,99
Metionina (M)	131,19	124,0	162,9	-0,23
Fenilalanina (F)	147,18	135,0	189,9	-1,13
Prolina (P)	97,12	90,0	112,7	0,45
Serina (S)	87,08	73,0	89,0	0,13
Treonina (T)	101,11	93,0	116,1	0,14
Triptofano (W)	186,21	163,0	227,8	-1,85
Tirosina (Y)	163,18	141,0	193,6	-0,94
Valina (V)	99,14	105,0	140,0	0,07

Referências: massa molar (LAMERS *et al.*, 2008), volume (DARBY; CREIGHTON, 1993) e (SYSTEM, 2017), hidrofobicidade (WIMLEY; WHITE, 1996).

Fonte – elaborado pelo autor.

O (χ^2) observado é calculado a partir dos dados das amostras:

$$\chi_{observado}^2 = \sum \frac{(o - e)^2}{e} \quad (3.2)$$

Onde o é a frequência de dados observada e e é a frequência esperada.

A tomada de decisão através do teste do χ^2 depende do χ^2 crítico tabelado, ou apenas χ_c^2 , que depende do número de graus de liberdade e do nível de significância adotado. O número de graus de liberdade é calculado pelo número de classes menos 1. O nível de significância representa a máxima probabilidade de erro que se tem ao rejeitar uma hipótese. O resultado do teste é interpretado comparando-se os valores do χ^2 calculado e o χ_c^2 :

- se χ^2 calculado $\geq \chi_c^2$ tabelado, as frequências observadas são diferentes das sequências esperadas;
- se χ^2 calculado $< \chi_c^2$ tabelado, as frequências observadas não são diferentes das sequências esperadas.

O χ_c^2 utilizado para os testes de validação das camadas de SASA neste trabalho é o valor de 7,815, pois o número de graus de liberdade é 3 (número de camadas de SASA menos um) e o nível de significância adotado foi 0,05 (Apêndice A).

Os testes estatísticos de Qui-quadrado e correlação de Pearson foram auxiliados pelo pacote Scikit de Python (v. 3.6).

3.6 Ampliação do Banco de Dados

A maioria das mutações depositadas no banco de dados, é, como esperado, desfavorável. Isso gera um problema de escassez de dados para se trabalhar na avaliação de mutações favoráveis, pois pode-se, por exemplo, realizar análises tendenciosas por serem baseadas em poucos dados, o que as torna pouco generalizáveis.

Para tratar este problema, trabalhou-se com o uso de mutações hipotéticas: para cada mutação contida no banco de dados onde um resíduo A é substituído por B (sentido direto), com $\Delta\Delta G = \Delta\Delta G_{A \rightarrow B}$, a mutação inversa $B \rightarrow A$ é integrada ao banco com valor $\Delta\Delta G_{B \rightarrow A} = -\Delta\Delta G_{A \rightarrow B}$.

Isso permite que cada mutação desfavorável dê origem a uma mutação favorável, e, com isso, o banco de dados passou a compreender 7210 mutações para a análise das mutações favoráveis.

4 RESULTADOS E DISCUSSÃO

O objetivo deste trabalho foi o de avaliar que características estruturais e de cada aminoácido tornam a estrutura tridimensional de proteínas mais estáveis. Para isso, foram analisadas mutações pontuais experimentais.

Dos seis descritores utilizados para avaliar o efeito de cada mutação, os caracteres estruturais possuem potencial para prever como o local da mutação influencia na estabilidade da proteína, enquanto que particularidades dos aminoácidos podem ser úteis para selecionar o resíduo que será usado na substituição.

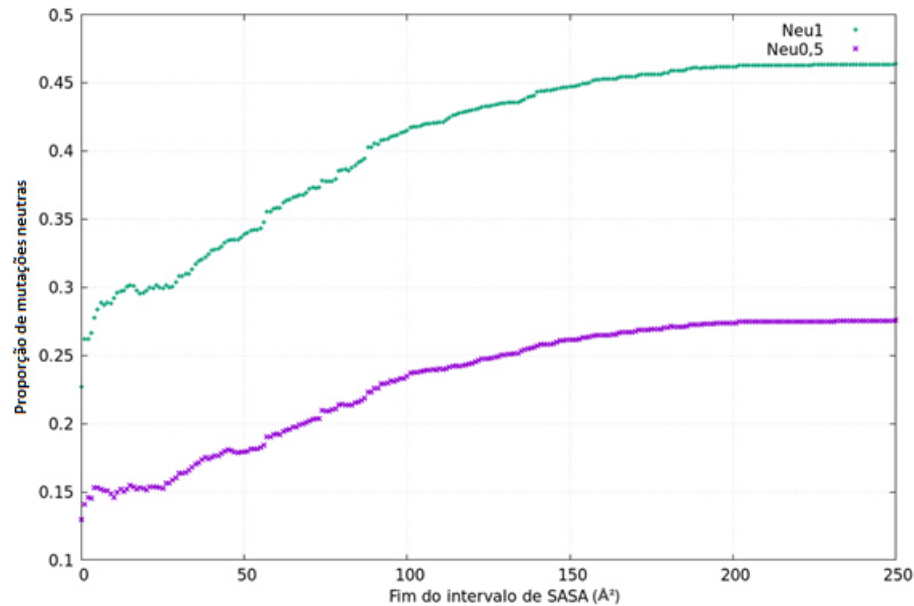
Dois características estruturais foram consideradas neste trabalho: tipo de estrutura secundária (ES) e a SASA. Quanto ao aminoácido, foram analisados os seguintes fatores em relação às diferenças entre os resíduos envolvidos: volume, hidrofobicidade, máxima acessibilidade ao solvente e massa molecular.

4.1 Classificação da Mudança de Estabilidade

O presente trabalho objetiva avaliar as características pertinentes a mutações neutras. É, portanto, impossível considerar a categoria proposta por Capriotti *et al.* (2005), Potapov *et al.* (2009), onde mutações são classificadas apenas em favoráveis e desfavoráveis. As demais classificações Khan e Vihinen (2010), Capriotti *et al.* (2008) foram comparadas pela razão de mutantes que integram a categoria *neutra* para diferentes intervalos de SASA. O comportamento da variação de proporção de mutações neutras é semelhante para as duas classificações (Figura 8).

A categoria Neu1 foi escolhida por conta da maior representação de mutações neutras, contendo 46% dos dados contra apenas 28% da categoria Neu0,5. Por isso, este trabalho adotará o critério de Khan e Vihinen (2010): mutações favoráveis possuem $\Delta\Delta G < -1,0$ kcal/mol; mutações neutras $-1,0$ kcal/mol $< \Delta\Delta G < 1,0$ kcal/mol; e mutações desfavoráveis $1,0$ kcal/mol $< \Delta\Delta G$.

Figura 8 – Razão de mutações neutras em intervalos de SASA



Proporção de mutantes neutros para cada valores de SASA de 0 a 250. Neu1 = categoria neutra definida por $-1,0 \text{ kcal/mol} < \Delta\Delta G < 1,0 \text{ kcal/mol}$; Neu0,5 = categoria neutra definida por $-0,5 \text{ kcal/mol} < \Delta\Delta G < 0,5 \text{ kcal/mol}$. Correlação de Pearson = 0,99.

Fonte – elaborada pelo autor

4.2 Descritores Estruturais

4.2.1 Acessibilidade ao Solvente

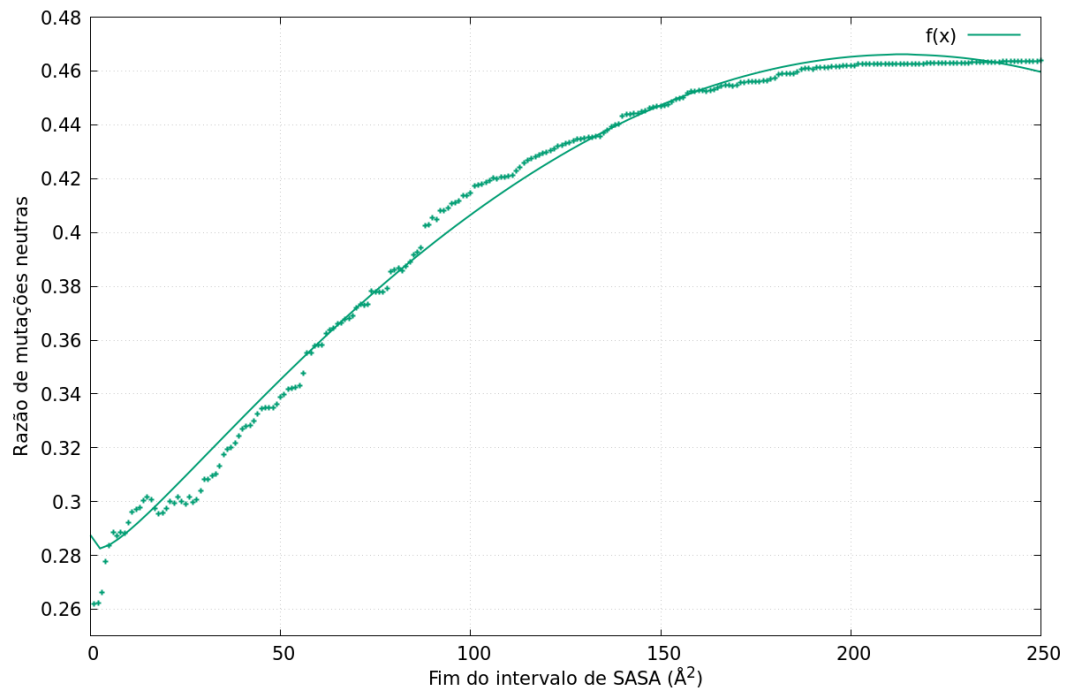
4.2.1.1 Mutações Neutras

As mutações neutras são predominantes para valores de SASA maiores (Figura 9). Isso ocorre porque mutações próximas ao centro hidrofóbico possuem mais chances de perturbar a estabilidade da molécula (SARABOJI *et al.*, 2005).

O número de mutações neutras cresce de maneira contínua entre os intervalos correspondentes a 32 \AA^2 e 100 \AA^2 , quando finalmente converge para um valor máximo de 46% (Figura 9). Isso permite dizer que, em geral, um sítio com $\text{SASA} = 150 \text{ \AA}^2$ deve ser preferencialmente selecionado para que a mutação seja neutra. Por outro lado, a razão de mutantes neutros encontrada para sítios com $\text{SASA} = 150 \text{ \AA}^2$ e 200 \AA^2 é efetivamente equivalente (40% de todo o intervalo de SASA analisado), variando apenas 3% em todo o intervalo. Isso indica que pouca ou nenhuma preferência pode ser estimada para sítios de $\text{SASA} \geq 150 \text{ \AA}^2$.

Aminoácidos são frequentemente categorizados de acordo com sua acessibilidade ao solvente (KHAN; VIHINEN, 2010; MASSO; VAISMAN, 2008; GIOLLO *et al.*, 2014;

Figura 9 – Razão de mutações neutras em intervalos de SASA



Cada valor de x indica intervalo de SASA de 0 a x.

$$f(x) = -0.00689946\sqrt{x} - 0.00000503479x^2 + 0.00238349x + 0.28755.$$

Fonte – elaborado pelo autor.

GUEROIS *et al.*, 2002; RULLER *et al.*, 2008). Embora os valores de acessibilidade ao solvente sejam contínuos, dividir os aminoácidos em algumas poucas categorias por sua SASA permite discretizar um domínio de valores em um conjunto de dados tratáveis. É, portanto, igualmente desejável que se dividam os aminoácidos em categorias de acordo com a proporção de mutações neutras, pois isso possibilita inferir parte da probabilidade de se obter uma mutação neutra para um dado sítio de acordo com sua acessibilidade ao solvente.

Tabela 6 – Proporção de mutações neutras em intervalos de SASA

Intervalo de SASA (Å ²)	Proporção de mutações neutras (%)
0 - 50	33,89
25 - 75	49,25
50 - 100	59,00
75 - 125	64,33
100 - 150	70,21
125 - 175	78,54
150 - 200	80,95
175 - 225	79,45
200 - 250	88,23

Fonte – elaborado pelo autor.

Para isso, compararam-se as proporções de mutações neutras para diferentes intervalos ao longo da amplitude de SASA de 0 \AA^2 a 250 \AA^2 (Tabela 6). Observa-se que sítios de mutação com SASA superior a 100 \AA^2 apresentaram mais de 70% de mutações neutras, variando de 70,21% de mutações neutras para o intervalo de 100 \AA^2 a 150 \AA^2 de SASA a 88,23% para o intervalo de 200 \AA^2 a 250 \AA^2 de SASA. As variações para os intervalos de 100 - 150, 125 - 175, 150 - 200, 175 - 225 e 200 - 250 são estatisticamente representativas da variação do intervalo 100 - 250 (Tabela 7), o que permite dizer que, para mutações neutras, o intervalo de SASA que varia de 100 \AA^2 a 250 \AA^2 constitui uma camada.

Tabela 7 – Intervalos de confiança de SASA

Intervalo de SASA	Proporção de mutações neutras (%)	Intervalo de confiança
100 - 150	70,21	3,28
125 - 175	78,54	5,27
150 - 200	80,95	7,16
175 - 225	79,45	11,37
200 - 250	88,23	19,78

Os intervalos de confiança se sobrepõem de forma que as proporções de mutações neutras possam ser consideradas um grupo unificado.

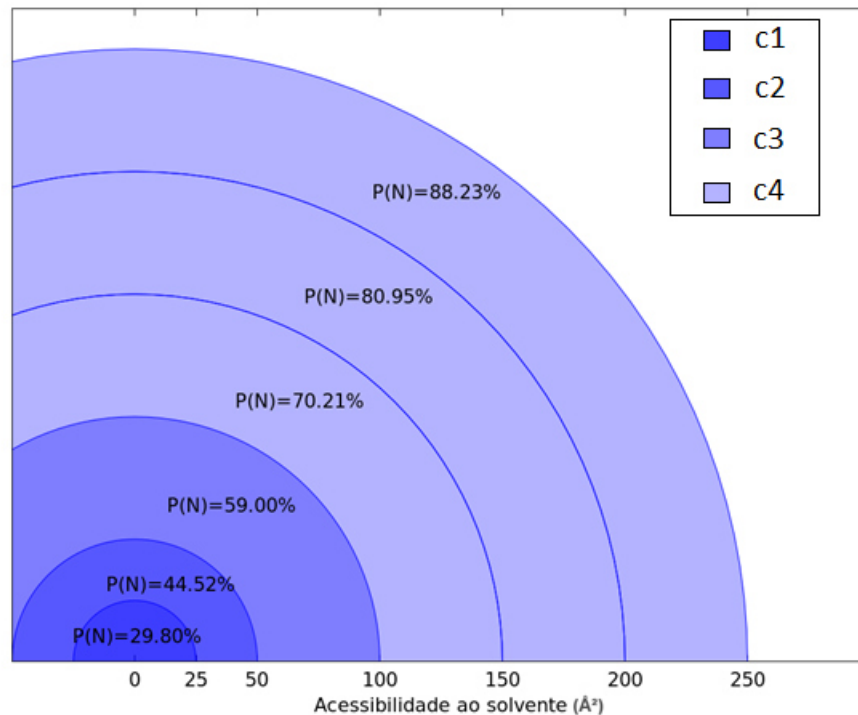
Fonte – elaborado pelo autor.

Valores muito próximos também foram observados nos intervalos de SASA a partir de 25 \AA^2 e abaixo de 100 \AA^2 , com número de mutações neutras entre 50% e 70%. Após a delimitação dessas três categorias quanto à SASA (0 \AA^2 a 25 \AA^2 , 25 \AA^2 a 100 \AA^2 e 100 \AA^2 a 250 \AA^2), foi decidido que a segunda categoria seria dividida em duas (delimitadas por $\text{SASA} = 50 \text{ \AA}^2$), pois abrangia uma larga amplitude de SASA e densidade de mutações (42% de todo banco de dados) (Figuras 10 e 11).

Um teste de validação foi feito para definir se as proporções de mutações neutras encontradas é um dado representativo dentro de cada camada. O teste consistiu na separação dos dados em dois subgrupos disjuntos formados aleatoriamente por 20% e 80% do banco de dados e a hipótese de igualdade estatística entre as proporções encontradas foi confirmada através do teste do qui-quadrado. Para isso, o subgrupo de 80% dos dados foi considerado o resultado esperado, enquanto que o subgrupo de 20% dos dados foi considerado como resultado observado.

O teste foi conduzido em triplicata e os valores de χ^2 observados variaram de 0,001616 a 0,009206 (Tabela 8), sendo menores que o valor do χ_c^2 tabelado (7,815), portanto os dados encontrados para os dois subconjuntos concordam e não apresentam desvios significativos

Figura 10 – Categorização quanto à acessibilidade ao solvente



P(N) = porcentagem de mutações neutras. A similaridade entre o número de mutações neutras sugere que a SASA pode ser dividida em 4 camadas: c1 = 0 - 25 Å², c2 = 25 - 50 Å², c3 = 50 - 100 Å², c4 = 100 - 250 Å².

Fonte – elaborado pelo autor.

quanto à frequência de mutações neutras em cada camada de SASA.

Tabela 8 – Qui-quadrado observado para mutações neutras

Intervalo de SASA	Teste 1		Teste 2		Teste 3	
	80%	20%	80%	20%	80%	20%
Camada enterrada	0.297275	0.300912	0.290349	0.32899	0.292562	0.318182
Camada semi-enterrada	0.446089	0.441667	0.438559	0.471074	0.446541	0.439655
Camada semi-exposta	0.592838	0.577381	0.585464	0.608939	0.584906	0.611111
Camada exposta	0.730942	0.759615	0.731651	0.754386	0.725275	0.789474
χ^2	0.001616		0.009200		0.009206	

Os valores na tabela indicam a razão de mutantes neutros em cada camada de SASA para conjuntos de 80% e 20% do banco de dados escolhidos aleatoriamente, em triplicata.

Camada enterrada: SASA de 0 Å² a 25 Å².

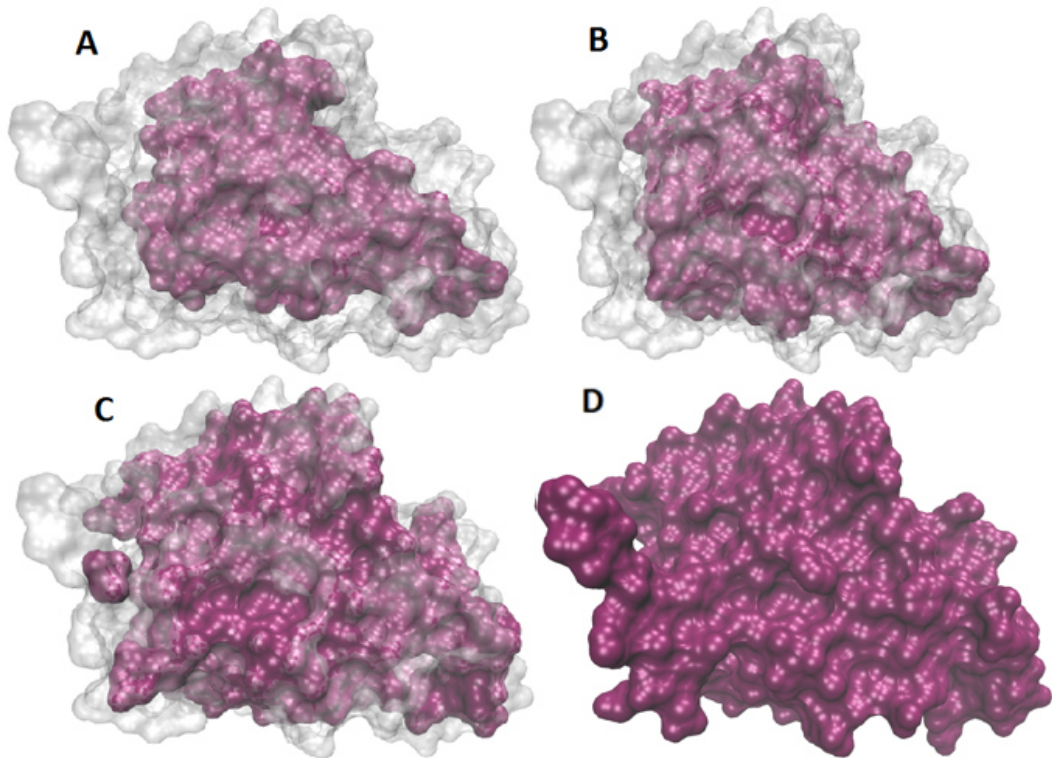
Camada semi-enterrada: SASA de 25 Å² a 50 Å².

Camada semi-exposta: SASA de 50 Å² a 100 Å².

Camada exposta: SASA de 100 Å² a 250 Å².

Fonte – elaborado pelo autor.

Figura 11 – Categorias quanto à acessibilidade ao solvente em cristal de uma proteína globular



Cristal de uma transferase de *Saccharomyces cerevisiae* (STRIEBECK *et al.*, 2013 (PDB ID: 3ZF8)). Em cada representação são pintados os resíduos com SASA menor que o valor máximo de cada categoria: A = até 25 Å²; B = até 50 Å²; C = até 100 Å²; D = até 250 Å².

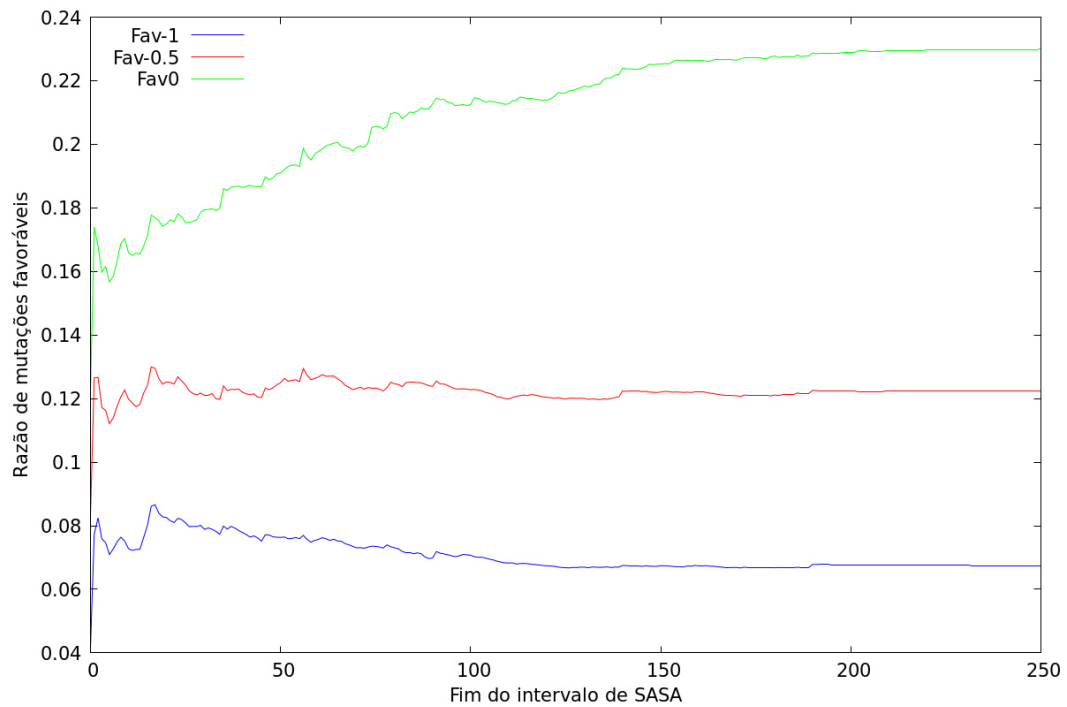
Fonte – elaborado pelo autor.

4.2.1.2 Ampliação do Banco de Dados

A relação entre mutações favoráveis e SASA foi comparada para as três categorizações consideradas inicialmente: $\Delta\Delta G \leq -1,0$ kcal/mol, $\Delta\Delta G \leq -0,5$ kcal/mol ou $\Delta\Delta G \leq 0$ kcal/mol (Figura 12). Diferente do resultado encontrado para as categorias neutras, não existe consenso entre o comportamento das categorias favoráveis. O comportamento para a categoria definida por $\Delta\Delta G \leq 0$ kcal/mol é particularmente discrepante em relação às outras, sendo predominantemente crescente. Tais discrepâncias podem sugerir que os dados não concordam por conta de uma baixa representatividade dos dados disponíveis para mutações favoráveis. De fato, 7% dos mutantes no banco de dados possuem $\Delta\Delta G$ menor que -1,0, e 12% dos mutantes no banco de dados possuem $\Delta\Delta G$ menor que -0,5, o que torna os grupos das mutações favoráveis pequenos, dificultando análises estatísticas dentro desses grupos. Para resolver esse problema, estratégias para tornar o grupo mais representável foram analisadas.

Capriotti *et al.* (2008), durante o treinamento do código do preditor de estabilidade I-Mutant3.0, propuseram ampliar o banco de dados para considerar que, além dos dados de

Figura 12 – Razão de mutações favoráveis em intervalos de SASA



Cada valor de x indica intervalo de SASA de 0 a x.

Fav-1 = categoria favorável definida por $\Delta\Delta G \leq -1$ kcal/mol;

Fav-0,5 = categoria favorável definida por $\Delta\Delta G \leq -0,5$ kcal/mol;

Fav0 = categoria favorável definida por $\Delta\Delta G \leq 0$ kcal/mol.

Fonte – elaborado pelo autor

mutações obtidos experimentalmente, o sentido inverso das mutações, ou seja, a mutação que transforma o mutante de volta na proteína original, pode ser incluído no banco de dados, apenas invertendo o sinal do $\Delta\Delta G$ associado à mutação. Resultados de outro trabalho (JIA *et al.*, 2015) indicam, ainda, que a adição de mutações inversas hipotéticas em um banco de dados durante o treinamento de preditores de estabilidade baseados em métodos estatísticos não apenas é válido, como também otimiza o poder de predição dos métodos.

Usar desse artifício resolve o problema da escassez de dados no grupo das mutações favoráveis por conta da adição dos dados inversos às mutações desfavoráveis, que representam 47% ($\Delta\Delta G \geq 1,0$ kcal/mol) todo o banco de dados.

O único requisito para que o banco de dados possa ser ampliado dessa forma é assumir que a Equação 4.1 seja verdadeira.

$$\Delta\Delta G_{XY} = -\Delta\Delta G_{YX} \quad (4.1)$$

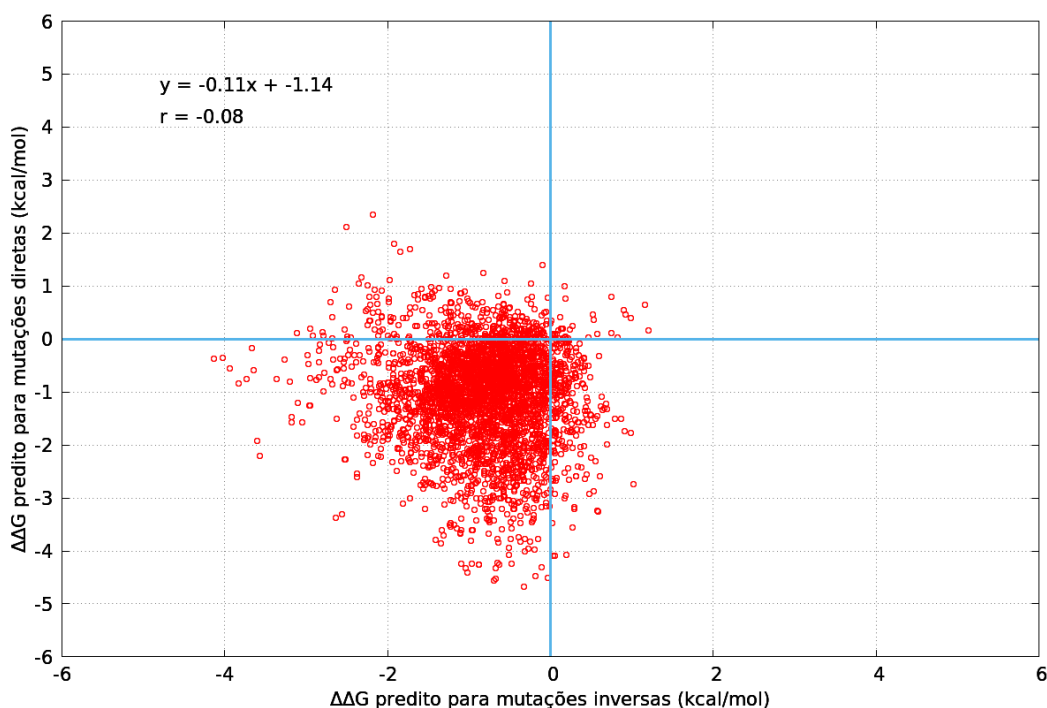
onde, $\Delta\Delta G_{XY}$ indica a diferença na variação de energia livre quando um resíduo X é mutado por um resíduo Y. Jia *et al.* (2015) e Li *et al.* (2012) discutem que a validade da igualdade é intuitiva,

uma vez que o $\Delta\Delta G$ é uma grandeza termodinâmica de estado, ou seja, que o seu cálculo depende apenas dos pontos inicial e final do processo de mutação (Equação 4.2).

$$\Delta\Delta G_{XY} = (\Delta G_Y - \Delta G_X) = -(\Delta G_X - \Delta G_Y) = -\Delta\Delta G_{YX} \quad (4.2)$$

Uma análise de correlação entre mutações diretas e inversas com um preditor de estabilidade (I-Mutant2.0 (THILTGEN; GOLDSTEIN, 2012)) mostrou resultado negativo (Figura 13), porém a acurácia dos métodos de predição é limitada por conta de aproximações feitas por cada método (THILTGEN; GOLDSTEIN, 2012). Por conta da relevância da base teórica que sustenta a Equação 4.1, o banco de dados foi ampliado, com a adição das mutações inversas hipotéticas, para seguir com análises dentro do grupo das mutações favoráveis.

Figura 13 – Correlação entre valores de $\Delta\Delta G$ preditos pelo I-Mutant2.0



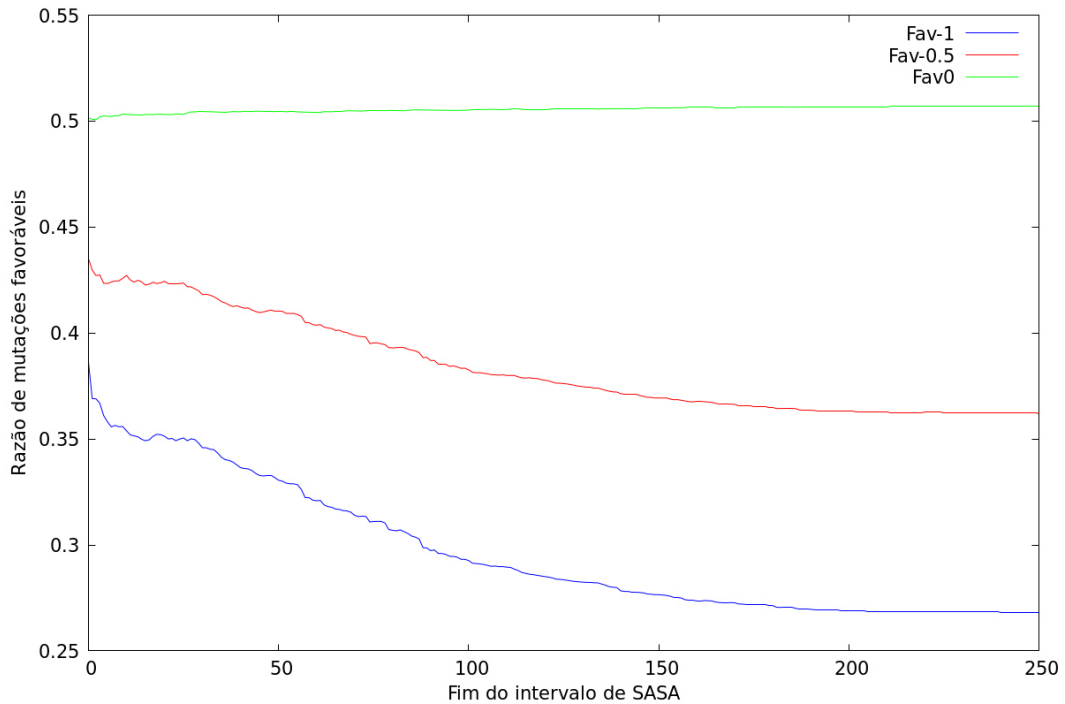
Mutações diretas se referem às mutações conforme descritas no banco de dados, enquanto que mutações inversas se referem às mutações hipotéticas geradas invertendo o sentido das mutações diretas. Os dados que concordam com a hipótese de inversão de mutações são localizados nas regiões onde o $\Delta\Delta G$ predito para mutação direta possui sinal contrário ao $\Delta\Delta G$ predito para a mutação inversa, correspondendo aos quadrantes superior esquerdo e inferior direito. Portanto o resultado não favorece a hipótese. O I-mutant2.0 é um campo de força baseado em aprendizado de máquina e, assim, possui performance limitada para predições de natureza diferente dos dados utilizados durante seu treinamento. Dessa forma, o campo de força provavelmente superestima características que contribuem para a atribuição de um $\Delta\Delta G$ desfavorável, já que esse tipo de mutação é predominantemente encontrado na literatura (CALIFORNIA, 2017) e, conseqüentemente, nos bancos de dados em que foi treinado (THILTGEN; GOLDSTEIN, 2012).

Fonte – elaborado pelo autor.

Para ilustrar a alteração no comportamento dos dados favoráveis, após ampliação do

banco de dados, as curvas para $\Delta\Delta G \leq -1,0$ kcal/mol, $\Delta\Delta G \leq -0,5$ kcal/mol e $\Delta\Delta G \leq 0$ kcal/mol são expostas na Figura 14. Observa-se que as categorias de mutações favoráveis Fav-0,5 e Fav-1 passam a se assemelhar após a ampliação do banco de dados e são considerados equivalentes.

Figura 14 – Razão de mutações favoráveis em intervalos de SASA



Cada valor de x indica intervalo de SASA de 0 a x.

Fav-1 = categoria favorável definida por $\Delta\Delta G \leq -1$ kcal/mol;

Fav-0,5 = categoria favorável definida por $\Delta\Delta G \leq -0,5$;

Fav0 = categoria favorável definida por $\Delta\Delta G \leq 0$.

Com a adição das mutações inversas hipotéticas no banco de dados, a quantidade de mutações com valores de $\Delta\Delta G$ positivos se torna idêntico ao grupo dos negativos, de modo que a curva Fav0 se torna praticamente constante, próxima a 50%. A correlação entre os dados obtidos para Fav-0,5 e Fav-1 apresentou coeficiente de Pearson de 0,99.

Fonte – elaborado pelo autor.

4.2.1.3 Mutações Favoráveis

Todos os intervalos exibidos na figura 9 possuem a porcentagem de mutações favoráveis abaixo de 35%, isso mostra que, até nas regiões de SASA onde a probabilidade de obtenção de mutações favoráveis é maior, não mais que um terço dos dados analisados é favorável.

A distribuição dos mutantes neutros e favoráveis nos intervalos analisados nas tabelas 9 e 6 possuem comportamento aparentemente complementar. De fato, uma vez que o banco de dados é ampliado, as mutações não neutras podem ser igualmente divididas no grupo das

Tabela 9 – Proporção de mutações favoráveis em intervalos de SASA

Intervalo de SASA (Å ²)	Proporção de mutações favoráveis (%)
0 - 50	33,05
25 - 75	25,37
50 - 100	20,50
75 - 125	17,84
100 - 150	14,90
125 - 175	10,73
150 - 200	9,52
175 - 225	10,27
200 - 250	5,88

Fonte – elaborado pelo autor.

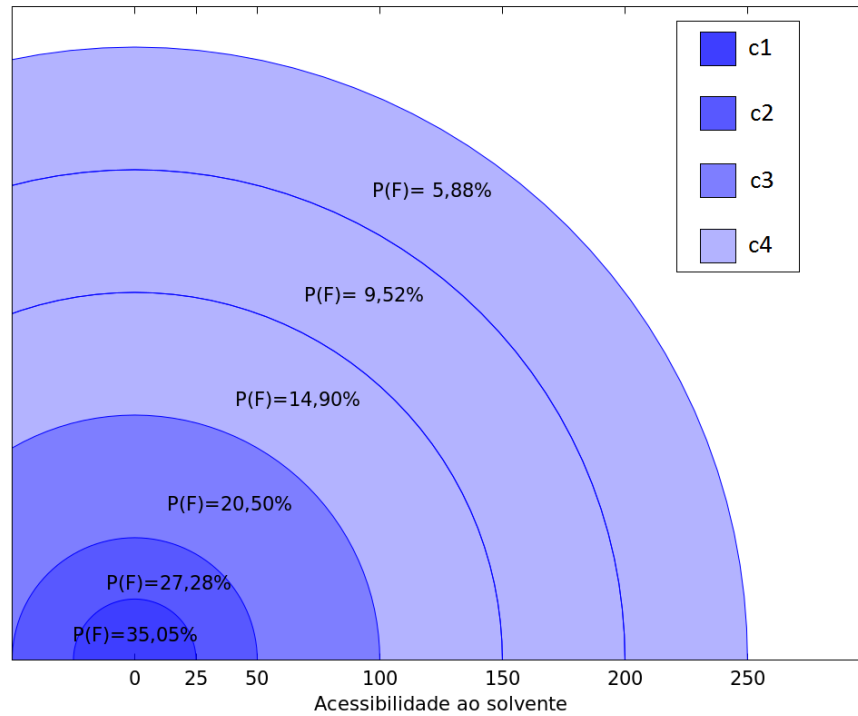
favoráveis e desfavoráveis, que possuem a mesma quantidade de dados e as mesmas características, considerando-se cada um dos descritores para essas mutações. Assim, tem-se que as porcentagens observadas para o grupo das mutações favoráveis são sempre a metade do complementar ao grupo das mutações neutras. Esse fato contribui para reforçar que mutações neutras e favoráveis possuem características antagônicas e que mutações favoráveis e desfavoráveis possuem características análogas, dependendo apenas do sentido em que a mutação ocorre (de um aminoácido A para outro B ou de B para A).

As proporções de mutações favoráveis foram calculadas para as camadas de SASA delimitadas durante a análise das mutações neutras (Figura 15). A complementaridade observada entre as mutações favoráveis e neutras torna que a definição das camadas seja igualmente representativa para ambos casos.

Como visto anteriormente, mutações que ocorrem entre valores de SASA de 100 Å² a 250 Å² são predominantemente neutras pois não perturbam diretamente o centro hidrofóbico e, portanto, impactos favoráveis ou desfavoráveis são menos comuns (SARABOJI *et al.*, 2005).

O teste estatístico do qui-quadrado foi novamente utilizado para avaliar a representatividade das proporções de mutações favoráveis encontradas para cada camada de SASA (Tabela 10). Os valores de χ^2 observados variaram de 0,008621 a 0,022051, sendo menores que o valor do χ_c^2 tabelado (7,815), portanto os dados encontrados para os dois subconjuntos concordam e não apresentam desvios significativos quanto à frequência de mutações neutras em cada camada de SASA.

Figura 15 – Frequência de mutações favoráveis por categoria de SASA



As porcentagens P(F) indicam a razão de mutações com impacto favorável na estabilidade para cada intervalo de SASA indicado. Diferentes tons de azul indicam as quatro categorias estabelecidas.

Fonte – elaborado pelo autor.

Tabela 10 – Qui-quadrado observado para mutações favoráveis

Intervalo de SASA	Teste 1		Teste 2		Teste 3	
	80%	20%	80%	20%	80%	20%
Camada enterrada	0.348122	0.363322	0.353878	0.339683	0.350703	0.351964
Camada semi-enterrada	0.266234	0.316794	0.276483	0.280992	0.27234	0.296748
Camada semi-exposta	0.206164	0.200521	0.203984	0.208763	0.213618	0.16763
Camada exposta	0.138322	0.105505	0.137255	0.104396	0.138158	0.101064
χ^2	0.018205		0.008621		0.022051	

Os valores na tabela indicam a razão de mutantes favoráveis em cada camada de SASA para conjuntos de 80% e 20% do banco de dados escolhidos aleatoriamente, em triplicata.

Camada enterrada: SASA de 0 Å² a 25 Å².

Camada semi-enterrada: SASA de 25 Å² a 50 Å².

Camada semi-exposta: SASA de 50 Å² a 100 Å².

Camada exposta: SASA de 100 Å² a 250 Å².

Fonte – elaborado pelo autor.

4.2.2 Estrutura secundária

Métodos de determinação de estrutura secundária a partir de um arquivo *Protein Data Bank* (PDB) se baseiam na detecção de padrões de distâncias entre carbonos- α , análise dos ângulos e comprimentos das ligações virtuais entre carbonos- α , análise de padrões de ligações

de hidrogênio, comparação com fragmentos de referência para cada elemento de estrutura secundária e quantificação da curvatura da cadeia principal (FRISHMAN; ARGOS, 1995). Neste trabalho, a estrutura secundária dos sítios de mutação foi determinada através do programa STRIDE (HEINIG; FRISHMAN, 2004), que utiliza padrões de ligações de hidrogênio (ângulo entre as ligações hidrogênio-doador e hidrogênio-receptor, e distância entre os átomos doador e receptor e propensões estatísticas dos ângulos formados entre dois aminoácidos vizinhos na cadeia principal de formarem elementos de estruturas secundária.

A análise de mutações neutras nos elementos de estrutura secundária mostra que aproximadamente metade das mutações eram neutras quando ocorreram em coil (52,94%) ou em hélice (49,61%), enquanto que em folhas- β , 35,52% das mutações foram neutras (Tabela 11). Quando comparadas as proporções de mutações neutras encontradas em elementos de ES para cada camada de SASA (c1, c2, c3 e c4), nota-se que o número de mutações ocorridas em elementos de ES definidos sofrem aumento à medida que o valor de SASA se distancia do centro, o que concorda com o que foi observado para as proporções de mutações neutras em geral para cada camada de SASA, o que indica que a ES pode não é capaz de causar grandes influências na obtenção de mutantes neutros.

A proporção de mutações neutras na camada c1 é semelhante para todos os elementos de ES (28,85% para coil, 28,40% para folha- β e 32,02% para hélice) e próximas a 29,80%, que corresponde a todas as mutações neutras na camada c1. Na camada c2, maior porcentagem de mutação neutra é observada em região de coil (49,60%) em comparação com as regiões de ES definida (38,06% em folha- β e 42,65% em hélice). Na camada c3 é onde parece ocorrer maior influência da ES pois as proporções se distanciam, com o menor valor observado em folha- β (46,50%) e maior valor para hélice (65,47%), porém ainda observa-se uma grande dependência da SASA, já que o número de mutações definidas na camada c3 é quase idêntico à proporção encontrada para a camada c3 (59,35% e 59,00%, respectivamente). Na camada c4, as proporções voltam a se assemelhar e variam de 72,50%, para coil, a 77,35%, para folha- β , mais uma vez concordando com a proporção de neutras em c4, igual a 73,63%.

Quanto a mutações favoráveis, a maior proporção é encontrada em folhas- β , onde 32,23% das mutações são favoráveis, enquanto as proporções de coil e hélice foram 23,52% e 25,19%, respectivamente. Os resultados observados para mutações favoráveis também concordam com as proporções observadas para as camadas de SASA. Na camada c1, as proporções encontradas foram de 35,57% para coil, 35,79% para folhas- β e 33,98% para hélice, enquanto

Tabela 11 – Proporção de mutações neutras por SASA e elemento de ES

Camada de SASA	Mutações neutras	Elemento de ES	Mutações neutras
Todas	46,00%	coil	52,94%
		folha- β	35,52%
		hélice	49,61%
		definido	43,08%
c1	29,80%	coil	28,85%
		folha- β	28,40%
		hélice	32,02%
		definido	30,03%
c2	44,51%	coil	49,60%
		folha- β	38,06%
		hélice	42,65%
		definido	40,12%
c3	59,00%	coil	58,27%
		folha- β	46,50%
		hélice	65,47%
		definido	59,35%
c4	73,63%	coil	72,50%
		folha- β	77,35%
		hélice	74,69%
		definido	75,34%

Fonte – elaborado pelo autor.

a proporção de favoráveis em c1 é de 35,09%. Em c2, a proporção de favoráveis é 27,74%, e as proporções para elementos de ES variam de 25,18%, para coil, a 30,96%, para folha- β . Em c3, as porcentagens encontradas para coil, definidas e para a camada independentemente de ES foram semelhantes (20,86%, 20,32% e 20,49%, respectivamente). As proporções em c4 foram de 13,74% para coil, 11,32% para folha- β e 12,65% para hélice, sem apresentar grandes distinções entre si ou quando comparadas à proporção de mutações favoráveis na camada c4, igual a 13,18%.

4.3 Descritores de identidade

Descritores baseados na identidade dos aminoácidos são importantes para determinar qual impacto uma substituição vai causar na estabilidade pois, uma vez selecionado o sítio de mutação, a escolha do aminoácido inserido deve decidir como a estabilidade será alterada. Como a estabilidade de uma proteína tem relação direta com a conformação tridimensional que é adotada pela molécula, as características mais relevantes identificadas em aminoácidos para prever alterações na estabilidade são relacionadas ao espaço que ocupam dentro da proteína

Tabela 12 – Proporção de mutações favoráveis por SASA e elemento de ES

Camada de SASA	Mutações favoráveis	Elemento de ES	Mutações favoráveis
Todas	27,00%	coil	23,52%
		folha- β	32,23%
		hélice	25,19%
		definido	28,45%
c1	35,09%	coil	35,57%
		folha- β	35,79%
		hélice	33,98%
		definido	34,98%
c2	27,74%	coil	25,18%
		folha- β	30,96%
		hélice	28,67%
		definido	29,93%
c3	20,49%	coil	20,86%
		folha- β	26,75%
		hélice	17,26%
		definido	20,32%
c4	13,18%	coil	13,74%
		folha- β	11,32%
		hélice	12,65%
		definido	12,32%

Fonte – elaborado pelo autor.

enovelada e a que tipo de interação fazem com outros resíduos e com o solvente.

Para exemplificar, são analisadas as mutações que ocorrem em locais onde a acessibilidade ao solvente é entre 150 e 250. Considerando o banco de dados original (sem ter sido ampliado), 164 mutações ocorrem no intervalo de 150 a 250 de SASA e, destas, 19 são desfavoráveis, 11 são favoráveis (Tabela 13) e as demais são neutras. Dentre as 19 mutações desfavoráveis, percebe-se que a maioria das mutações (14) ocorreram com a remoção de lisina (K) ou arginina (R), que são resíduos cuja cadeia lateral é ionizável e geralmente apresentam carga positiva em pH biológico. Dentre as mutações desfavoráveis ocorridas em K e R, somente uma possui cadeia lateral com capacidade de se tornar positivamente carregada. Somado a isso, dentre as restantes 5 mutações desfavoráveis, com SASA entre 150 e 250, e que não ocorrem removendo resíduos com carga positiva, duas ocorrem com a substituição de ácido glutâmico (E), que possui carga negativa, por resíduos sem carga. Esses dados podem sugerir que em ambientes com SASA entre 150 e 250, mutações que substituem resíduos carregados por resíduos não carregados devem ser evitadas pois possuem maior chance de resultar em uma estrutura que não é capaz de se enovelar corretamente.

Analogamente, analisando as mutações favoráveis com SASA entre 150 e 250,

Tabela 13 – Mutações favoráveis e desfavoráveis com SASA entre 150 e 250

Proteína	Mutação	$\Delta\Delta G$	SASA	Proteína	Mutação	$\Delta\Delta G$	SASA
1AMQ	K68E	Desfavorável	161,1	1LZ1	R21A	Favorável	156,3
1AMQ	K68M	Desfavorável	161,1	1LZ1	R21G	Favorável	156,3
1CYC	K73F	Desfavorável	163,0	1RN1	D49H	Favorável	158,1
1CYC	K73M	Desfavorável	163,0	1STN	K116G	Favorável	171,1
1CYC	K73W	Desfavorável	163,0	1HFZ	K114N	Favorável	181,0
1CYC	K73Y	Desfavorável	163,0	3SSI	M73D	Favorável	185,6
2CI2	I56A	Desfavorável	164,3	5CRO	Y26C	Favorável	189,1
1STN	Y85G	Desfavorável	165,8	5CRO	Y26D	Favorável	189,1
1C9O	R3A	Desfavorável	166,5	5CRO	Y26H	Favorável	189,1
1AMQ	F6A	Desfavorável	166,8	5CRO	Y26L	Favorável	189,1
1TUP	R248Q	Desfavorável	168,6	5CRO	Y26Q	Favorável	189,1
1AMQ	E7A	Desfavorável	182,1				
1VQB	E40C	Desfavorável	186,5				
1YCC	K73H	Desfavorável	189,6				
1YCC	K73I	Desfavorável	189,6				
1YCC	K73V	Desfavorável	189,6				
1YCC	K73W	Desfavorável	189,6				
1IOB	R4Q	Desfavorável	204,8				
2CI2	K72N	Desfavorável	208,4				

As proteínas selvagens são identificadas através de seu código PDB. Mutações desfavoráveis possuem $\Delta\Delta G \geq 1,0$ kcal/mol. Mutações favoráveis possuem $\Delta\Delta G \leq -1,0$ kcal/mol. Dados considerando o banco de dados original, sem ter sido ampliado.

Fonte – elaborado pelo autor.

percebe-se, primeiramente, que nenhuma mutação favorável ocorre com SASA acima de 190. Das 11 mutações favoráveis com SASA entre 150 e 250, nas 6 que ocorrem nas proteínas 5CRO e 3SSI há substituição do resíduo original por um aminoácido menos hidrofóbico (Tabela 5). De fato, as substituições feitas em 5CRO mostram cinco mutações que ocorrem no mesmo sítio e possuem um impacto favorável na estabilidade. Por outro lado, as demais 5 mutações favoráveis que ocorrem em sítios de SASA entre 150 e 250 geram aumento de hidrofobicidade, porém consistem na substituição de um resíduo carregado por um resíduo não carregado, o que pode poderia, por exemplo, ser explicado pelo fato de os resíduos removidos serem um ponto fraco na estabilidade mas foram conservados evolutivamente porque atribuem outra função à proteína, como integrar o sítio ativo ou participar da interação com um ligante.

Assim, percebe-se que descritores baseados na identidade de aminoácidos são fundamentais para o entendimento de qual mutação deve ser conduzida para se atingir determinado objetivo.

Os descritores que dependem da identidade dos aminoácidos foram avaliados através da classificação dos mutantes em termos de SASA ou impacto na estabilidade. Após atribuído o valor que cada aminoácido possui para cada descritor quantitativo, a variação absoluta do valor

do parâmetro, quando comparados selvagem e mutante é calculada e a distribuição desses valores é analisada em gráficos de diagramas de caixa.

4.3.1 *Volume e Massa molecular*

Massa e volume são descritores que estão relacionados ao espaço ocupado pela proteína. O espaço que a cadeia lateral de cada resíduo ocupa é diretamente contemplado pelo volume dos aminoácidos, porém, um trabalho que utilizou descritores de aminoácidos como *input* para o treinamento de preditores de estabilidade estatísticos define a massa molecular como um descritor com possível contribuição para a performance dos preditores (JIA *et al.*, 2015). Portanto, a massa molecular também foi considerada inicialmente.

Foi utilizada a massa molecular média de cada resíduo de aminoácido (CANTÚ *et al.*, 2008). O espaço ocupado pela cadeia lateral de cada resíduo depende das interações que os átomos da cadeia lateral fazem, por isso, cálculo exato do volume depende da estrutura e da localização do aminoácido (ZAMYATNIN, 1972). Portanto, o volume tabelado é uma aproximação e, para este trabalho, foi analisado a partir de duas referências (ZAMYATNIN, 1972; DARBY; CREIGHTON, 1993).

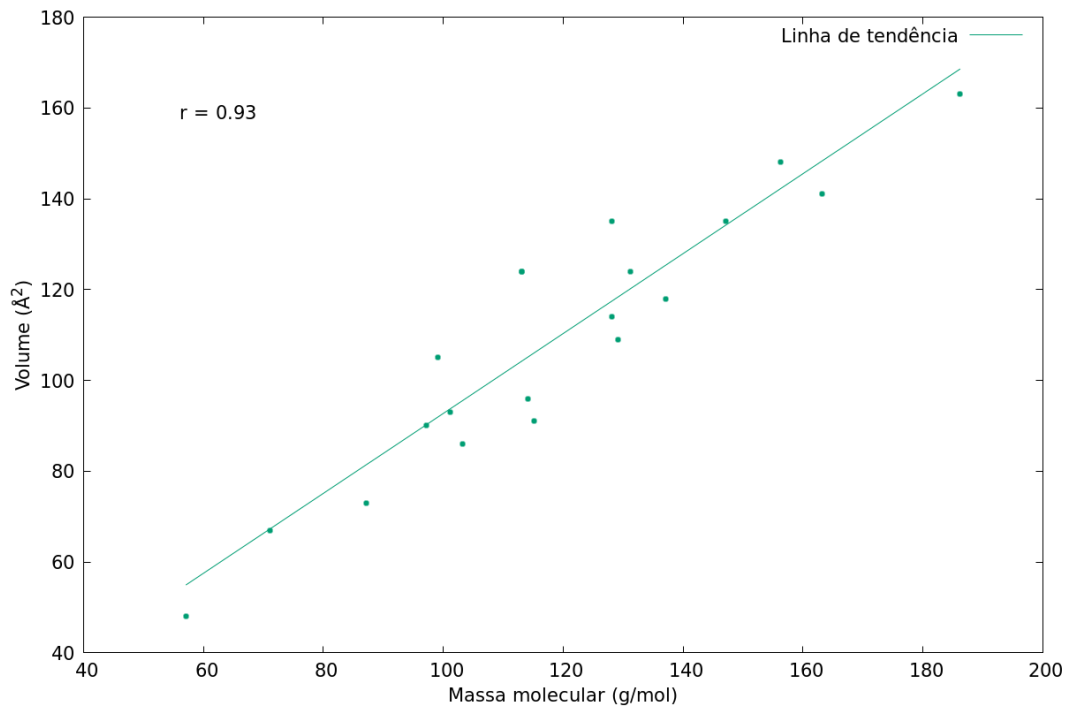
A utilização de massa molecular se mostrou equivalente ao volume (Figura 16), assim, foram consideradas apenas as distribuições baseadas no volume dos aminoácidos.

A diferença entre o volume do aminoácido mutante e o volume do aminoácido selvagem (Δ volume) foi calculada para cada mutação do banco de dados e sua distribuição foi analisada separadamente para os grupos das mutações favoráveis, não-favoráveis (neutras e desfavoráveis) e neutras (Figuras 17 e 18). Como o grupo que contém as mutações não favoráveis é composto pelas mutações desfavoráveis e neutras, ele deve mostrar sempre resultados intermediários entre a distribuição de mutações neutras e de mutações favoráveis, já que, como os valores de Δ volume analisados são absolutos, as distribuições para favoráveis e desfavoráveis são idênticas.

Valores maiores de Δ volume (Δ volume = 100 Å³, por exemplo) indicam que houve substituição entre dois aminoácidos que possuem maior diferença de volume, sem levar em conta se houve troca de um resíduo menor por um maior ou um maior por um menor.

Na figura 17, não são observadas variações significativas quando comparados os grupos. Na figura 18, contudo, percebe-se que a distribuição dos dados entre neutras e favoráveis é distinta, com a mediana do grupo das neutras em 28 Å³ e a mediada do grupo das favoráveis

Figura 16 – Correlação entre volume e massa molecular



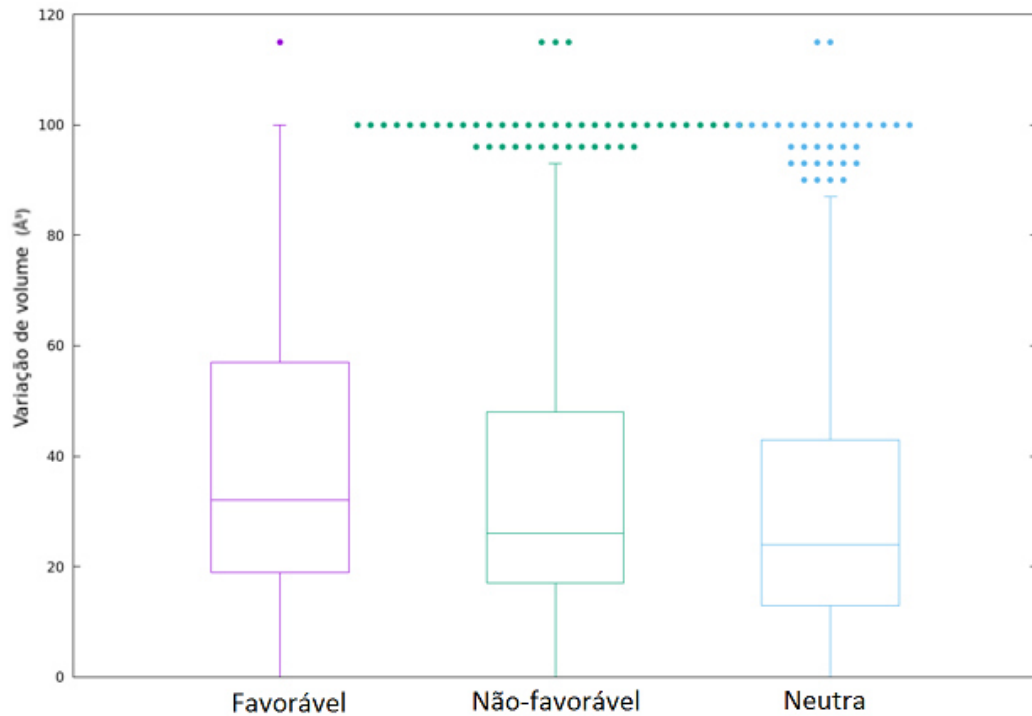
Os volumes dos aminoácidos foram retirados de Darby e Creighton (1993). O valor de r se refere ao coeficiente de correlação de Pearson entre volume e massa molecular.

Fonte – elaborado pelo autor.

em 52 \AA^3 , indicando que mais mutações neutras ocorrem quando a substituição ocorre entre aminoácidos de volume mais parecido. Apesar da predominância de valores de Δvolume abaixo de 50 \AA^3 , são também observadas mutações neutras para substituições de aminoácidos com grandes diferenças de volume. A análise do Δvolume combinado à SASA pode elucidar se existem padrões na distribuição desses dados.

Dentre mutações neutras, sítios com SASA entre 0 \AA^2 e 25 \AA^2 (enterrados na proteína) compreendem a região mais crítica para a predição de mutações neutras, devido à maior proporção de mutações favoráveis. Comparando-se mutações enterradas neutras e não neutras (Figuras 19 e 20), nota-se que mutações neutras estão concentradas em valores menores de Δvolume , de modo que, na figura 20, 75% das mutações neutras possuem Δvolume de até 50 \AA^3 , enquanto que 75% das mutações não-neutras variam até 80 \AA^3 . De fato, substituições enterradas em que o resíduo mutante possui maior volume que o resíduo selvagem, facilmente geram grande modificação espacial no centro da proteína, produzindo uma mutação não-neutra (DEHOUCK *et al.*, 2011). Mutações neutras são mais prováveis de ocorrer quando os resíduos mutante e selvagem possuem volume semelhante.

Figura 17 – Distribuição de Δ volume por tipo de impacto na estabilidade



Os volumes dos aminoácidos foram retirados de Darby e Creighton (1993).

Fonte – elaborado pelo autor.

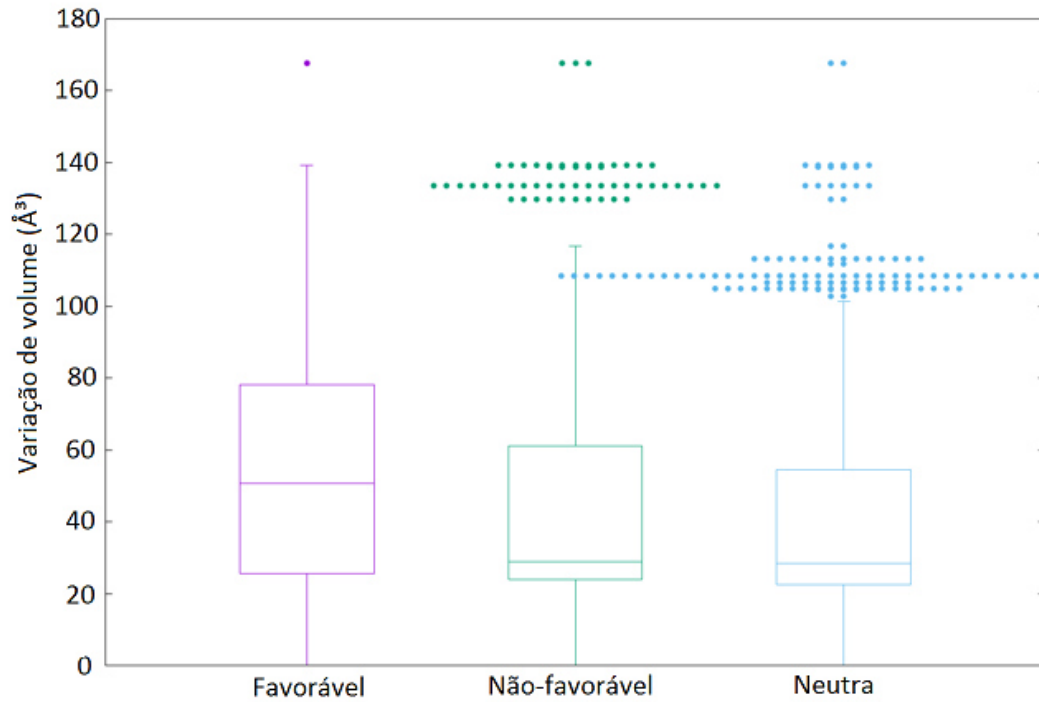
4.3.2 Hidrofobicidade

A hidrofobicidade de cada aminoácido é determinada experimentalmente como a energia livre de transferência de cada aminoácido de um ambiente aquoso para a interface de uma membrana dupla de fosfolípidios (WHITE, 2011).

A hidrofobicidade de aminoácidos é uma característica que está diretamente ligada à predisposição de um resíduo de se encontrar naturalmente em determinada região da proteína, no centro hidrofóbico ou na superfície (BIOSTRUCTURES, 2016). Apesar da hidrofobicidade ter estreita relação com a localização de resíduos na proteína, pouca diferença foi encontrada quando analisadas as distribuições de Δ hidrofobicidade em mutações favoráveis, não-favoráveis e neutras (Figura 21).

Levando em conta apenas mutações enterradas, valores mais baixos de variação de hidrofobicidade são observados dentre mutações neutras, onde 50% dos dados são iguais ou menores que 0,5 kcal/mol e o maior valor Δ hidrofobicidade é 2,2 kcal/mol. Enquanto isso, o grupo das mutações não-neutras mostrou-se com mediana por volta de 0,9 kcal/mol e valor máximo de Δ hidrofobicidade igual a 3,09 kcal/mol.

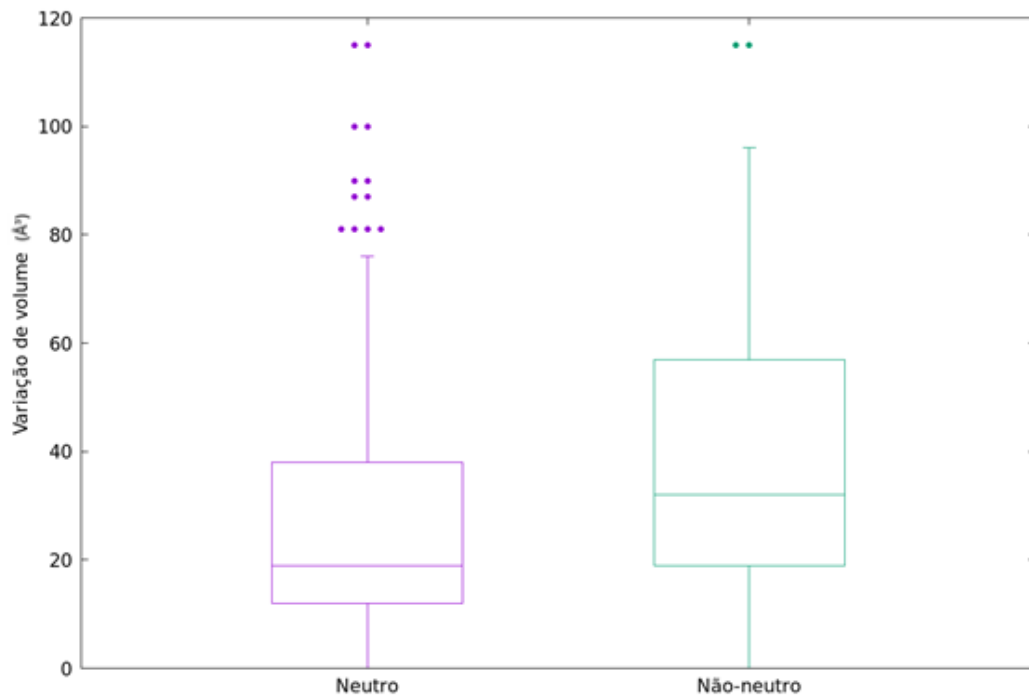
Figura 18 – Distribuição de Δ volume por tipo de impacto na estabilidade



Os volumes dos aminoácidos foram retirados de Zamyatnin (1972)

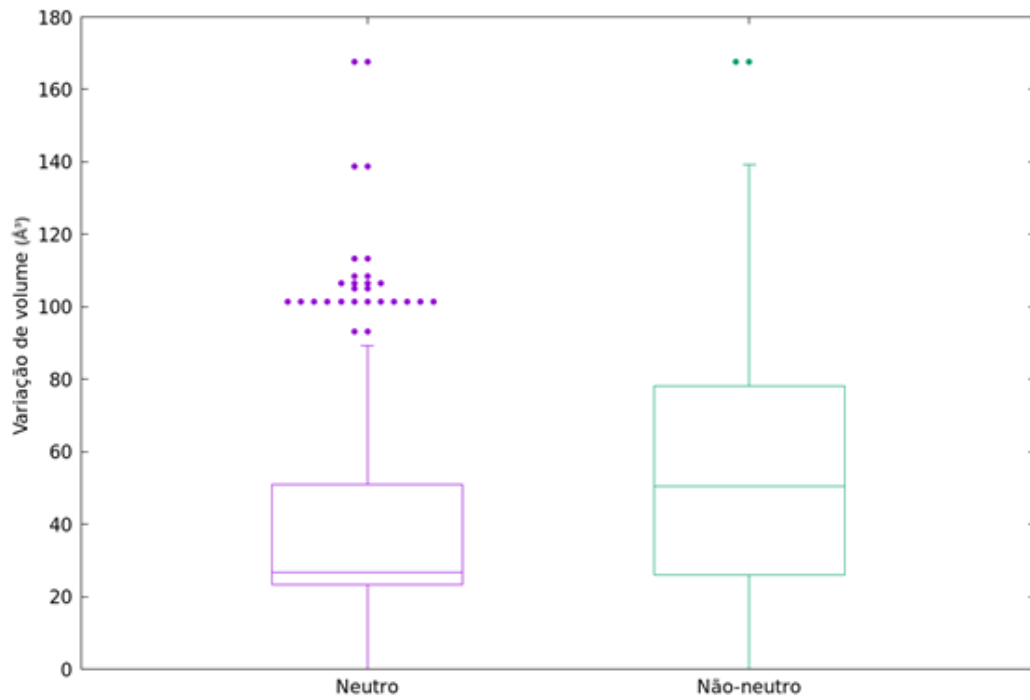
Fonte – elaborado pelo autor.

Figura 19 – Δ volume de mutações enterradas



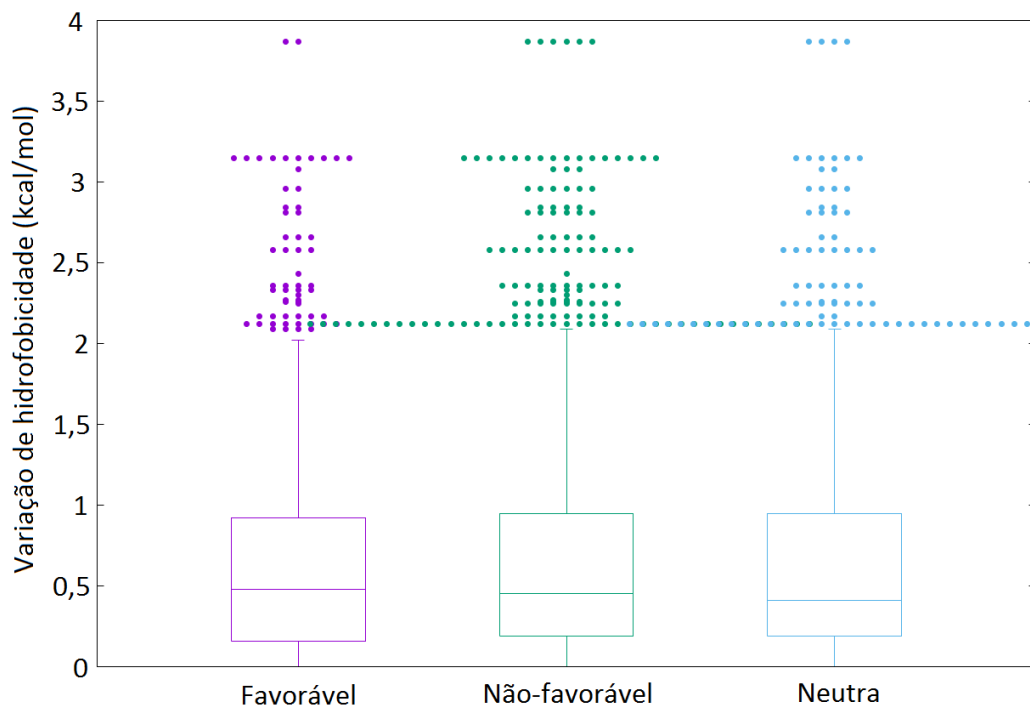
Os volumes dos aminoácidos foram retirados de Darby e Creighton (1993)

Fonte – elaborado pelo autor.

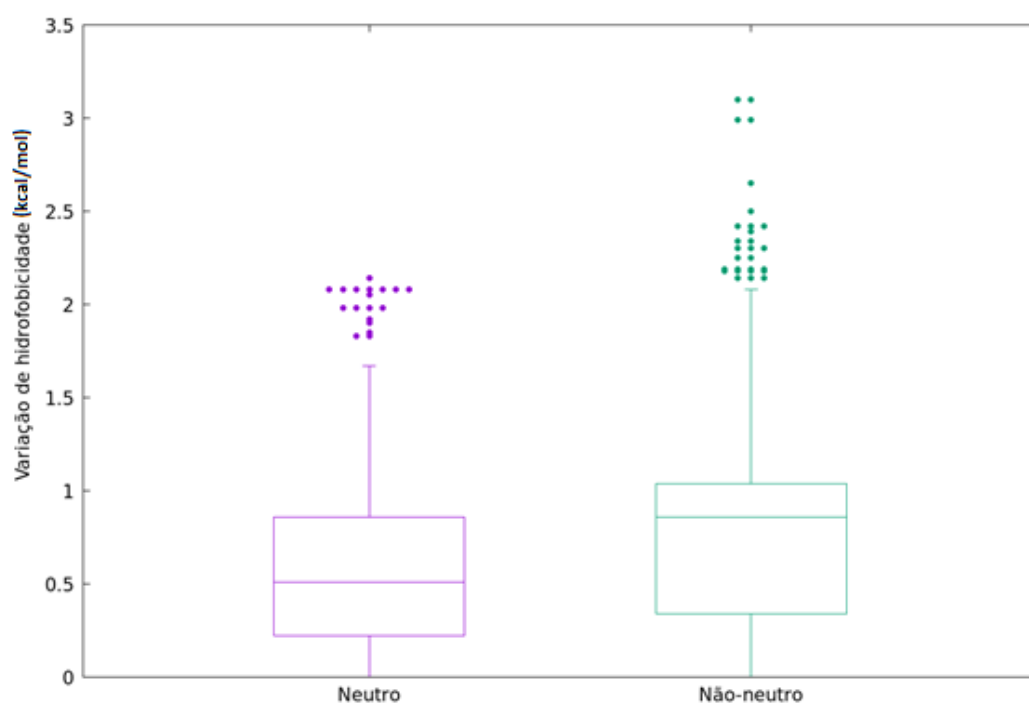
Figura 20 – Δ volume de mutações enterradas

Os volumes dos aminoácidos foram retirados de Zamyatnin (1972)

Fonte – elaborado pelo autor.

Figura 21 – Distribuição de Δ hidrofobicidade por tipo de impacto na estabilidade

Fonte – elaborado pelo autor.

Figura 22 – Δ Hidrofobicidade de mutações enterradas

Fonte – elaborado pelo autor.

5 CONCLUSÕES E PERSPECTIVAS

Dentre os descritores estruturais analisados, a área superficial acessível ao solvente foi o parâmetro que apresentou maior influência na determinação da variação de estabilidade de proteínas submetidas a mutações. Camadas de SASA se mostraram representativas e permitiram uma análise clara da influência de combinações de SASA com os demais parâmetros analisados. A estrutura secundária não apresentou contribuição relevante para a predição do impacto que mutações causam na estabilidade. Dentre os descritores de identidade dos aminoácidos, massa molar e volume se mostraram equivalentes e relevantes. A hidrofobicidade não obteve resultados muito relevantes na diferenciação de mutações neutras e favoráveis, sua influência se restringiu apenas quando analisadas mutações enterradas.

REFERÊNCIAS

- ANDRADE, M. A.; O'DONOGHUE, S. I.; ROST, B. Adaptation of protein surfaces to subcellular location. **Journal of Molecular Biology**, Academic Press Limited (London), v. 276, p. 517–525, 1998.
- BAKER, E. N.; HUBBARD, R. E. Hydrogen bonding in globular proteins. **Progress in Biophysics and Molecular Biology**, Pergamon Press Ltd, v. 44, p. 97–179, 1984.
- BALCÃO, V. M.; VILA, M. M. D. C. Structural and functional stabilization of protein entities: *state-of-the-art*. **Advanced Drug Delivery Reviews**, Elsevier Science, Ltd., v. 93, p. 25–41, 2015.
- BERLAND, K.; COOPER, V. R.; LEE, K.; SCHRÖDER, E.; THONHAUSER, T.; HYLDGAARD, P.; LUNDQVIST, B. I. van der waals force in density functional theory: a review of the vdw-df methd. **Reports on Progress in Physics**, IOP Publishing Ltd., v. 78, n. 066501, p. –, 2015.
- BERMAN, H. M.; WESTBROOK, J.; FENG, Z.; GILLILAND, G.; BHAT, T. N.; WEISSIG, H.; SHINDYALOV, I. N.; BOURNE, P. E. The protein data bank. **Nucleic Acids Research**, v. 28, p. 235–242, 2000.
- BIEDERMANN, F.; NAU, W. M.; SCHNEIDER, H. J. The hydrophobic effect revisited — studies with supramolecular complexes imply high-energy water as a noncovalent driving force. **Angewandte Chemie International Edition**, Wiley-VCH, v. 53, p. 11158–11171, 2014.
- BIOSTRUCTURES, S. **Structural bioinformatics: practical guide**. 2016. Disponível em: <<http://www.proteinstructures.com/index.html>>. Acesso em: 16 jun. 2017.
- BOMMARIUS, A. S.; BROERING, J. M.; CHAPARRO-RIGGERS, J. F.; POLIZZI, K. M. High-throughput screening for enhanced protein stability. **Current Opinion in Biotechnology**, v. 17, p. 606–610, 2006.
- BRYANT, C. M.; MCCLEMENTS, D. J. Molecular basis of protein functionality with special consideration to cold-set gels derived from heat-denatured whey. **Trends in Food Science and Technology**, Elsevier Science, Ltd., v. 9, p. 143–151, 1998.
- BUTT, H. J.; CAPPELLA, B.; KAPPL, M. Force measurements with the atomic force microscope: Technique, interpretation and applications. **Surface Science Reports**, Elsevier Science, Ltd., v. 59, p. 1–152, 2005.
- BUYEL, J. F.; TWYMAN, R. M.; FISCHER, R. Very-large-scale production of antibodies in plants: The biologization of manufacturing. **Biotechnology Advances**, Elsevier Science, Ltd., v. 35, p. 458–465, 2017.
- CALIFORNIA, T. R. of the University of. **Macromolecular Modeling and Design Benchmarks and Benchmark Captures**. 2017. Disponível em: <<https://guybrush.ucsf.edu/benchmarks/index>>. Acesso em: 28 jun. 2017.
- CANTÚ, M. D.; CARRILHO, E.; WULFF, N. A.; PALMA, M. S. Sequenciamento de peptídeos usando espectrometria de massas: um guia prático. **Química Nova**, v. 31, n. 3, p. 669–675, 2008.

- CAPRIOTTI, E.; FARISELLI, P.; CASADIO, R. I-mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. **Nucleic Acids Research**, Oxford University Press, v. 33, p. W306–W310, 2005.
- CAPRIOTTI, E.; FARISELLI, P.; ROSSI, I.; CASADIO, R. A three-state prediction of single point mutations on protein stability changes. **BMC Bioinformatics**, Italian Society of Bioinformatics, v. 9, p. (Suppl 2):S6, 2008.
- CHANDLER, D. Interfaces and the driving force of hydrophobic assembly. **Nature**, Nature Publishing Group, v. 437, n. 29, p. 640–647, 2005.
- CHEN, H.; ZHOU, H. X. Prediction of interface residues in protein-protein complexes by a consensus neural network method: test against nmr data. **Proteins: Structure, Function and Bioinformatics**, Wiley-Liss, Inc., v. 61, p. 21–35, 2005.
- CHEN, H.; ZHOU, H. X. Prediction of solvent accessibility and sites of deleterious mutations from protein sequence. **Nucleic Acids Research**, Oxford university Press, v. 33, n. 10, p. 3193–3199, 2005.
- DARBY, N. J.; CREIGHTON, T. E. **Protein structure**. [S.l.]: IRL Press, 1993.
- DAY, A. **The Source of Stability in Proteins**. 1996. Disponível em: <<http://www.cryst.bbk.ac.uk/PPS2/projects/day/TDayDiss/index.html>>. Acesso em: 28 abr. 2017.
- DEHOUCQ, Y.; KWASIGROCH, J. M.; GILIS, D.; ROOMAN, M. Popmusic 2.1: a web server for the estimation of protein stability changes upon mutation and sequence optimality. **BMC Bioinformatics**, PubMed Central, v. 12, n. 151, 2011.
- DEMAIN, A. L. The business of biotechnology. **Industrial Biotechnology**, Gen Publishing, Inc., v. 3, n. 3, p. 269–283, 2007.
- DEUTSCH, C.; KRISHNAMOORTHY, B. Four-body scoring function for mutagenesis. **Bioinformatics**, v. 23, n. 22, p. 3009–3015, 2007.
- DEY, T. B.; KUMAR, A.; BANERJEE, R.; CHANDNA, P.; KUHAD, R. C. Improvement of microbial alpha-amylase stability: strategic approaches. **Process Biochemistry**, Elsevier Science, Ltd., v. 51, p. 1380–1390, 2016.
- DILL, K. A. Dominant forces in protein folding. **Biochemistry**, American Chemical Society, v. 29, n. 31, p. 7133–7155, 1990.
- DILL, K. A.; MACCALLUM, J. L. The protein-folding problem, 50 years on. **Science**, American Association for the Advancement of Science, v. 338, p. 1042–1046, 2012.
- EISENHABER, F.; LIJNZAAD, P.; ARGOS, P.; SANDER, C.; SCHARF, M. The double cubic lattice method: efficient approaches to numerical integration of surface area and volume and to dot surface contouring of molecular assemblies. **Journal of Computational Chemistry**, John Wiley and Sons, Inc., v. 16, n. 3, p. 273–284, 1995.
- FRISHMAN, D.; ARGOS, P. Knowledge-based protein secondary structure assignment. **Proteins: Structure, Function and Genetics**, Wiley-Liss, Inc., v. 23, p. 566–579, 1995.

- GIOLLO, M.; MARTIN, A. J. M.; WALSH, I.; FERRARI, C.; TOSATTO, S. C. E. Neemo: a method using residue interaction networks to improve prediction of protein stability upon mutation. **BMC Genomics**, BioMed Central Ltd., v. 15, p. (Suppl 4):S7, 2014.
- GREAVES, T. L.; DRUMMOND, C. J. Solvent nanostructure, the solvophobic effect and amphiphile self-assembly in ionic liquids. **Chemical Society Reviews**, The Royal Society of Chemistry, v. 42, n. 3, p. 1096–1120, 2013.
- GUEROIS, R.; NIELSEN, J. E.; SERRANO, L. Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. **Journal of Molecular Biology**, Elsevier Science, Ltd., v. 320, p. 369–387, 2002.
- HARTL, F. U.; HAYER-HARTL, M. Converging concepts of protein folding *in vitro* and *in vivo*. **Nature Structural and Molecular Biology**, Nature America, Inc., v. 16, n. 6, p. 574–581, 2009.
- HEINIG, M.; FRISHMAN, D. Stride: a web server for secondary structure assignment from known atomic coordinates of proteins. **Nucleic Acids Research**, v. 32, p. W500–2, 2004.
- HENDSCH, Z. S.; TIDOR, B. Do salt bridges stabilize proteins? a continuum electrostatic analysis. **Protein Science**, The Protein Society, v. 3, p. 211–226, 1994.
- HOFFMANN, F.; VANCEA, I.; KAMAT, S. G.; STRODEL, B. Protein structure prediction: assembly of secondary structure elements by basin-hopping. **ChemPhysChem**, Wiley-VCH, v. 15, p. 3378–3390, 2014.
- HUYGHUES-DESPOINTES, B. M. P.; PACE, C. N.; ENGLANDER, S. W.; SCHOLTZ, J. M. **Protein structure, stability and folding**. [S.l.]: Humana Press, Inc., 2001. v. 168.
- ITO, Y.; IKEUCHI, A.; IMAMURA, C. Advanced evolutionary molecular engineering to produce thermostable cellulase by using a small but efficient library. **Protein Engineering, Design & Selection**, v. 26, n. 1, p. 73–79, 2013.
- JIA, L.; YARLAGADDA, R.; REED, C. C. Structure based thermostability prediction models for protein single point mutations with machine learning tools. **PLoS ONE**, v. 10, n. 9, p. e0138022., 2015.
- KABSCH, W.; SANDER, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. **Biopolymers**, John Wiley and Sons, Inc., v. 22, p. 2577–2637, 1983.
- KELOGG, E. H.; LEAVER-FAY, A.; BAKER, D. Role of conformational sampling in computing mutation-induced changes in protein structure and stability. **Proteins: structure, function and bioinformatics**, Wiley-Liss, Inc., v. 79, p. 830–838, 2011.
- KHAN, S.; VIHINEN, M. Performance of protein stability predictors. **Human Mutation**, Wiley-Liss, Inc., v. 31, n. 6, p. 675–684, 2010.
- KUMAR, R. Role of naturally occurring osmolytes in protein folding and stability. **Archives of Biochemistry and Biophysics**, Elsevier Science, Ltd., v. 491, p. 1–6, 2009.
- LAMERS, S. L.; SALEMI, M.; MCGRATH, M. S.; FOGEL, G. B. Prediction of r5, x4, and r5x4 hiv-1 coreceptor usage with evolved neural networks. **IEEE/ACM Transactions on Computational Biology and Bioinformatics**, v. 5, n. 2, p. 291–300, 2008.

- LEE, J.; BLABER, M. The interaction between thermodynamic stability and buried free cysteines in regulating the functional half-life of fibroblast growth factor-1. **Journal of Molecular Biology**, Elsevier, Ltd., v. 393, p. 113–127, 2009.
- LEITE, F. L.; BUENO, C. C.; RÓZ, A. L. d.; ZIEMATH, E. C.; JUNIOR, O. N. O. Theoretical models for surface forces and adhesion and their measurement using atomic force microscopy. **International Journal of Molecular Sciences**, v. 13, p. 12773–12856, 2012.
- LI, Y.; ZHANG, J.; TAI, D.; MIDDAUGH, C. R.; ZHANG, Y.; FANG, J. Prots: a fragment based protein thermo-stability potential. **Proteins**, Wiley Periodicals, Inc., v. 80, n. 1, p. 81–92, 2012.
- MÁRQUEZ-CHAMORRO, A. E.; ASECIO-CORTEZ, G.; SANTIESTEBAN-TOCA, C. E.; AQQUILAR-RUIZ, J. S. Soft computing methods for the prediction of protein tertiary structures: A survey. **Applied Soft Computing**, Elsevier Science, Ltd., v. 35, p. 398–410, 2015.
- MASSO, M.; VAISMAN, I. I. Accurate prediction of stability changes in protein mutants by combining machine learning with structure based computational mutagenesis. **Bioinformatics**, Oxford University Press, v. 24, n. 18, p. 2002–2009, 2008.
- MITCHELL, L. A.; CAI, Y.; TAYLOR, M.; NORONHA, A. M.; CHUANG, J.; DAI, L.; BOEKE, K. D. Multi change isothermal mutagenesis: a new strategy for multiple site-directed mutations in plasmid dna. **ACS Synthetic Biology**, American Chemical Society, 2012. No prelo.
- OCHS, R. S. Thermodynamics and spontaneity. **Journal of Chemical Education**, v. 73, n. 10, p. 952–954, 1996.
- OLIVEIRA, L. G. d.; MANTOVANI, S. M. Transformações biológicas: contribuições e perspectivas. **Química Nova**, v. 32, n. 3, p. 742–756, 2009.
- PACE, C. N.; GRIMSLEY, G. R.; SCHOLTZ, J. M. **Protein stability**. [S.l.]: John Wiley and Sons, Ltd., 2009.
- PACE, C. N.; SHIRLEY, B. A.; MCNUTT, M.; GAJIWALA, A. Forces contributing to the conformational stability of proteins. **The FASEB Journal**, FASEB, v. 10, n. 1, p. 75–83, 1996.
- PARANÁ, U. F. do. **Qui quadrado**. 2011. Disponível em: <<http://www.ufpa.br/dicas/biome/bioqui.htm>>. Acesso em: 15 jun. 2017.
- POTAPOV, V.; COHEN, M.; SCHREIBER, G. Assessing computational methods for predicting protein stability upon mutation: good on average but not in the details. **Protein Engineering, Design and Selection**, Oxford University Press, v. 22, n. 9, p. 553–560, 2009.
- REECE, J. B.; URRY, L. A.; CAIN, M. L.; WASSERMANN, S. A.; MINORSKY, P. V.; JACKSON, R. B. **Biologia de Campbell**. [S.l.]: Editora Artmed, 2015.
- REETZ, M. T. Laboratory evolution of stereoselective enzymes: A prolific source of catalysts for asymmetric reactions. **Angewandte Chemie International Edition**, Wiley-VCH, v. 50, p. 138–174, 2011.
- REETZ, M. T. **Directed evolution of selective enzymes: catalysts for organic chemistry and biotechnology. First edition**. [S.l.]: Wiley-VCH, 2017.

- REVA, B.; ANTIPIN, Y.; SANDER, C. Predicting the functional impact of protein mutations: application to cancer genomics. **Nucleic Acids Research**, Oxford University Press, v. 39, n. 17, p. e118, 2011.
- RICHMOND, T. J. Solvent accessible surface area and excluded volume in proteins: Analytical equations for overlapping spheres and implications for the hydrophobic effect. **Journal of Molecular Biology**, Academic Press, Inc. (London), v. 178, p. 63–89, 1984.
- RULLER, R.; DELIBERTO, L.; FERREIRA, T. L.; WARD, R. J. Thermostable variants of the recombinant xylanase a from *Bacillus subtilis* produced by directed evolution show reduced heat capacity changes. **Proteins**, Wiley-Liss, Inc., v. 70, p. 1280–1293, 2008.
- SANTOS, M. M. d. **Utilização de mutagênese aleatória para obtenção da lipase de *Burkholderia cepacia* com variação nas propriedades catalíticas**. Dissertação (Mestrado) — Universidade Estadual do Paraná, Curitiba, 2005.
- SARABOJI, K.; GROMIHA, M. M.; PONNUSWAMY, M. N. Relative importance of secondary structure and solvent accessibility to the stability of protein mutants. **Computational Biology and Chemistry**, Elsevier Science, Ltd., v. 29, p. 25–35, 2005.
- SHAHEEN, R.; AMASHA, H.; ALJAMALI, M. Protein solvent accessibility prediction systems. **American Journal of Biomedical and Life Sciences**, Science Publishing Group, v. 3, n. 2-3, p. 21–24, 2015.
- SIDDIQUI, K. S. Some like it hot, some like it cold: Temperature dependent biotechnological applications and improvements in extremophilic enzymes. **Biotechnology Advances**, Elsevier Science, Ltd., v. 33, p. 1912–1922, 2015.
- SILVA, M. C. M. d.; FIGUEIRA, E. L. Z.; Sá, M. F. G. d. **A metodologia de DNA shuffling na produção de diversidade gênica**. [S.l.]: Embrapa Recursos Genéticos, 2003.
- STAHLKE, D. **Mercado global de enzimas industriais**. [S.l.]: Associação Brasileira da Indústria e Comércio de Ingredientes e Aditivos para Alimentos, 2015.
- STRIEBECK, A.; ROBINSON, D. A.; SCHUETTELKOPF, A. W.; AALTEN, D. M. F. V. Yeast *mnn9* is both a priming glycosyltransferase and an allosteric activator of mannan biosynthesis. **Open Biology**, v. 3, n. 130022, p. –, 2013 (PDB ID: 3ZF8).
- STUDER, R. A.; CHRISTIN, P. A.; WILLIAMS, M. A.; ORENGO, C. A. Stability-activity tradeoffs constrain the adaptive evolution of rubisco. **PNAS**, v. 111, n. 6, p. 2223–2228, 2014.
- STURM, N. **Amino acids/Proteins**. 2015. Disponível em: <http://www.nbs.csudh.edu/chemistry/faculty/nsturm/CHE450/04_AminoAcidsProteins.htm>. Acesso em: 2 jun. 2017.
- SYSTEM, T. I. I. **Amino acid abbreviations, characteristics, volume and hydropathy index**. 2017. Disponível em: <http://www.imgt.org/IMGTeducation/Aide-memoire/_UK/aminoacids/abbreviation.html>. Acesso em: 16 jun. 2017.
- THILTGEN, G.; GOLDSTEIN, R. A. Assessing predictors of changes in protein stability upon mutation using self-consistency. **PLoS ONE**, v. 7, n. 10, p. e46084, 2012.
- TOUW, W. G.; BAAKMAN, C.; BLACK, J.; BEEK, T. A. H. t.; KRIEGER, E.; JOOSTEN, R. P.; VRIEND, G. A series of pdb related databases for everyday needs. **Nucleic Acids Research**, v. 43, p. D364–D368, 2015.

WANG, W. Instability, stabilization, and formulation of liquid protein. **International Journal of Pharmaceutics**, Elsevier Science, Ltd., v. 185, p. 129–188, 1999.

WHITE, S. **Experimentally determined hydrophobicity scales**. 2011. Disponível em: <http://blanco.biomol.uci.edu/hydrophobicity_scales.html>. Acesso em: 16 jun. 2017.

WIMLEY, W. C.; WHITE, S. H. Experimentally determined hydrophobicity scale for proteins at membrane interfaces. **Nature Structural Biology**, v. 3, n. 10, p. 842–848, 1996.

ZAMYATNIN, A. A. Protein volume in solution. **Progress in Biophysics and Molecular Biology**, v. 24, p. 107–123, 1972.

ANEXO A – TABELA DE QUI-QUADRADO

Tabela 14 – χ^2 tabelado

GL	0,99	0,90	0,50	0,30	0,20	0,10	0,05	0,02	0,01	0,001
01	0,0002	0,016	0,455	1,074	1,642	2,706	3,841	5,412	6,635	10,827
02	0,020	0,211	1,386	2,408	3,219	4,605	5,991	7,824	9,210	13,815
03	0,115	0,584	2,366	3,665	4,642	6,251	7,815	9,837	11,345	16,266
04	0,297	1,064	3,357	4,878	5,989	7,779	9,488	11,668	13,277	18,467
05	0,554	1,610	4,351	6,064	7,289	9,236	11,070	13,388	15,080	20,515
06	0,872	2,204	5,348	7,231	8,558	10,645	12,592	15,033	16,812	22,457
07	1,239	2,833	6,346	8,383	9,803	12,017	14,067	16,622	18,475	24,322
08	1,646	3,490	7,344	9,524	11,030	13,362	15,507	18,168	20,090	26,125
09	2,088	4,168	8,343	10,656	12,242	14,684	16,919	19,679	21,666	27,877
10	2,558	4,865	9,342	11,781	13,442	15,987	18,307	21,161	23,209	29,588
11	3,053	5,578	10,341	12,899	14,631	17,275	19,675	22,618	24,725	31,264
12	3,571	6,304	11,340	14,011	15,812	18,549	21,026	24,054	26,217	32,909
13	4,107	7,042	12,340	15,119	16,985	19,812	22,362	25,472	27,688	34,528
14	4,660	7,790	13,339	16,222	18,151	21,064	23,685	26,873	29,141	36,123
15	5,229	8,547	14,339	17,322	19,311	22,307	24,996	28,259	30,578	37,697
16	5,812	9,312	15,338	18,418	20,465	23,542	26,296	29,633	32,000	39,252
17	6,408	10,085	16,338	19,511	21,615	24,769	27,587	30,995	33,409	40,790
18	7,015	10,865	17,338	20,601	22,760	25,989	28,869	32,346	34,805	42,312
19	7,633	11,651	18,338	21,689	23,900	27,204	30,144	33,687	36,191	43,820
20	8,260	12,443	19,337	22,775	25,038	28,412	31,410	35,020	37,566	45,315

No topo: GL (grau de liberdade) e níveis de significância. Em negrito, valores de χ_c^2 para o nível de significância mais comumente adotado (0,05).

Fonte – (PARANÁ, 2011).