

UNIVERSIDADE FEDERAL DE MINAS GERAIS
PROGRAMA INTERUNIDADES DE PÓS-GRADUAÇÃO EM BIOINFORMÁTICA

LETÍCIA XAVIER SILVA CANTÃO

**Modelagem molecular aliada ao aprendizado de máquina na
busca por assinaturas de resistência a herbicidas em Acetolactato
sintases (ALSs)**

Belo Horizonte
2021

UNIVERSIDADE FEDERAL DE MINAS GERAIS
PROGRAMA INTERUNIDADES DE PÓS-GRADUAÇÃO EM BIOINFORMÁTICA

LETÍCIA XAVIER SILVA CANTÃO

**Modelagem molecular aliada ao aprendizado de máquina na
busca por assinaturas de resistência a herbicidas em Acetolactato
sintases (ALSs)**

Dissertação apresentada ao Programa Interunidades de Pós-Graduação em Bioinformática da Universidade Federal de Minas Gerais, como requisito à obtenção do título de Mestre em Bioinformática.

Orientador: Raquel Cardoso de Melo Minardi

Coorientador: Leonardo Henrique França de Lima

Belo Horizonte

2021

- 043 Cantão, Letícia Xavier Silva.
Modelagem molecular aliada ao aprendizado de máquina na busca por assinaturas de resistência a herbicidas em Acetolactato sintases (ALSs) [manuscrito] / Letícia Xavier Silva Cantão. – 2021.
79 f. : il. ; 29,5 cm.
- Orientador: Raquel Cardoso de Melo Minardi. Coorientador: Leonardo Henrique França de Lima.
Dissertação (mestrado) – Universidade Federal de Minas Gerais, Instituto de Ciências Biológicas. Programa Interunidades de Pós-Graduação em Bioinformática.
1. Biologia Computacional. 2. Acetolactato Sintase. 3. Resistência a Herbicidas. 4. Aprendizado de Máquina. I. Minardi, Raquel Cardoso de Melo. II. Lima, Leonardo Henrique França de. III. Universidade Federal de Minas Gerais. Instituto de Ciências Biológicas. IV. Título.

CDU: 573:004



UNIVERSIDADE FEDERAL DE MINAS GERAIS
Instituto de Ciências Biológicas
Programa Interunidades de Pós-Graduação em Bioinformática da UFMG

ATA DE DEFESA DE DISSERTAÇÃO

LETÍCIA XAVIER SILVA CANTÃO

Às nove horas do dia **13 de setembro de 2021**, reuniu-se, através de videoconferência, a Comissão Examinadora de Dissertação, indicada pelo Colegiado do Programa, para julgar, em exame final, o trabalho intitulado: "**Modelagem molecular aliada ao aprendizado de máquina na busca por assinaturas de resistência a herbicidas em Acetolactato sintases (ALSs)**", requisito para obtenção do grau de Mestre em **Bioinformática**. Abrindo a sessão, a Presidente da Comissão, **Dra. Raquel Cardoso de Melo Minardi**, após dar a conhecer aos presentes o teor das Normas Regulamentares do Trabalho Final, passou a palavra à candidata, para apresentação de seu trabalho. Seguiu-se a arguição pelos Examinadores, com a respectiva defesa da candidata. Logo após, a Comissão se reuniu, sem a presença da candidata e do público, para julgamento e expedição de resultado final. Foram atribuídas as seguintes indicações:

Professor(a)/Pesquisador(a)	Instituição	Indicação
Dra. Raquel Cardoso de Melo Minardi	Universidade Federal de Minas Gerais	Aprovada
Dr. Leonardo Henrique Franca de Lima	Universidade Federal de São João Del-Rei	Aprovada
Dr. Ernesto Raúl Caffarena	FIOCRUZ	Aprovada
Dra. Rafaela Salgado Ferreira	Universidade Federal de Minas Gerais	Aprovada
Dr. Maurílio Fernandes de Oliveira	Embrapa	Aprovada

Pelas indicações, a candidata foi considerada: **Aprovada**

O resultado final foi comunicado publicamente à candidata pela Presidente da Comissão. Nada mais havendo a tratar, a Presidente encerrou a reunião e lavrou a presente ATA, que será assinada por todos os membros participantes da Comissão Examinadora.

Belo Horizonte, 13 de setembro de 2021.



Documento assinado eletronicamente por **Ernesto Raul Caffarena, Usuário Externo**, em 13/09/2021, às 11:57, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Rafaela Salgado Ferreira, Professora do Magistério Superior**, em 13/09/2021, às 11:57, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Maurilio Fernandes de Oliveira, Usuário Externo**, em 13/09/2021, às 14:02, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **0954367** e o código CRC **46BFDAE3**.

Referência: Processo nº 23072.247554/2021-63

SEI nº 0954367

Agradecimentos

Gostaria de agradecer primeiramente a Deus por mais essa conquista em minha vida. Por me manter saudável, cheia de vida e cercada por pessoas maravilhosas.

Agradeço de forma carinhosa aos meus pais. Em especial a minha querida mãe, que foi e sempre será minha inspiração e meu orgulho. Sempre esteve ao meu lado nos dias bons e dias ruins, me incentivando e me mostrando que eu era capaz e que sempre é possível se superar. Agradeço minha irmã, que sempre me amou e me apoiou em tudo.

Meu amado esposo, cheio de sabedoria e inteligência, que me compreendeu, me entendeu e me ajudou a superar cada limitação. Que comemorou comigo todas as minhas conquistas e esteve comigo me incentivando e me dando amor e carinho durante esses 12 anos juntos. Obrigada Arthur por ser meu parceiro de vida. Eu te amo.

Gostaria de agradecer pelo apoio e paciência de meus amigos, especialmente Isabella e Samara entre os muitos que amo. Mesmo tão ocupada e sem muito tempo para saídas e bate-papo estiveram ao meu lado, nunca desistiram da nossa amizade, pelo contrário só a fortalecemos. Aos meus avós, tios, primos e sogros meu eterno obrigada, vocês foram minha alegria em dias de encontros e risadas.

Agradeço aos colegas do laboratório LBS, que me ajudaram durante minha caminhada. Rafael, Diego entre outros que me socorreram algumas vezes, muito obrigada. Minha colega Luana, que se tornou uma amiga para a vida, muito obrigada por ser minha dupla.

Agradeço de forma carinhosa aos meus queridos professores Dra. Raquel Minardi e Dr. Leonardo França, pela oportunidade de realizar esse trabalho ao lado de mentes brilhantes. Obrigada pelo carinho, prontidão, paciência e obrigada pelos ensinamentos acadêmicos, mas também pelos ensinamentos para a vida. Obrigada pela dedicação e esforço para que este trabalho fosse concluído. Agradeço de forma especial pela colaboração enriquecedora do professor Dr. Paulo Ricardo Batista.

Agradeço também às agências de fomento à pesquisa: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) e Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq); ao Programa de Pós-graduação em Bioinformática da UFMG, principalmente Sheila e Tiago por serem pessoas extremamente dedicadas, pacientes e talentosas. O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

Resumo

Aproximadamente 60% dos pesticidas utilizados em culturas de plantas são herbicidas destinados à eliminação de plantas daninhas. A enzima ALS ou acetohidroxiácido sintase (AHAS; EC 2.2.1.6) é alvo de inibição por cinco classes de herbicidas e está envolvida na via de biossíntese de aminoácidos de cadeia ramificada (valina, leucina e isoleucina). A exposição contínua desses herbicidas às culturas levou à evolução de plantas daninhas de biótipos resistentes a herbicidas. Esses biótipos apresentam geralmente resistência local-alvo, sendo documentadas mutações pontuais em várias espécies. O desenvolvimento de culturas de plantas resistentes aos herbicidas atuais (contribuindo para que apenas as plantas daninhas sejam atingidas) e o desenvolvimento de novos herbicidas se tornou extremamente necessário. Em busca por assinaturas de resistência a dois tipos de herbicidas (sulfoniluréias e imidazolinonas) em ALS, aqui atrelamos o aprendizado de máquina a dados de modelagem molecular de enzimas com a presença do inibidor, com e sem mutações. Simulações de dinâmica molecular, e outras técnicas de bioinformática estrutural foram levadas à técnicas de seleção de atributos utilizadas em aprendizado de máquina de forma a melhor discernir atributos da ALSs que separam resistentes e susceptíveis. Os resultados sugerem que o mecanismo de ganho ou não de resistência aos herbicidas com as mutações está ligado a mudanças tanto na dinâmica, rede de contatos e perfil energético no complexo proteína-ligante. Para a sulfoniluréia (SU) as alterações nestes atributos sugerem uma maior restauração da componente competitiva do inibidor na proteína em relação à imidazolinona (IMI), em consonância com a própria maior componente inibitória relatada na literatura para as SU comparado às IMI. Em enzimas com o inibidor imidazolinona, a resistência parece ter relação direta com modificações alostéricas, modificando estruturalmente a região dos cofatores. Em enzimas com o inibidor sulfonilureia o padrão de resistência sugere uma forte relação com a perda de afinidade pelo ligante. Os resultados aqui obtidos podem contribuir na elucidação de novos caminhos para a temática sustentável de plantas daninhas, culturas e herbicidas. Estes mesmos resultados também demonstram que o uso de aprendizado de máquina para encontrar padrões em meio a uma diversidade de dados retornados pela modelagem e dinâmica molecular é uma estratégia prontamente aplicável e eficaz.

Palavras-chave: acetolactato sintases; ALS; resistência a herbicidas; aprendizado de máquina; seleção de atributos.

Abstract

Approximately 60% of pesticides used on plant crops are herbicides aimed at eliminating weeds. The enzyme ALS or acetohydroxy acid synthase (AHAS; EC 2.2.1.6) is targeted for inhibition by five classes of herbicides and is involved in the pathway of branched-chain amino acid biosynthesis (valine, leucine and isoleucine). Continuous exposure of these herbicides to crops has led to the evolution of weeds of herbicide resistant biotypes. These biotypes generally present target-site resistance, with point mutations being documented in several species. The development of plant crops resistant to current herbicides (contributing to only weeds being affected) and the development of new herbicides has become extremely necessary. In search of resistance signatures to two types of herbicides (sulfonylureas and imidazolinones) in ALS, here we link machine learning to molecular modeling data of enzymes with the presence of the inhibitor, with and without mutations. Molecular dynamics simulations, and other structural bioinformatics techniques were taken to attribute selection techniques used in machine learning in order to better discern attributes of ALSs that separate resistant and susceptible. The results suggest that the mechanism of gain or not of resistance to herbicides with mutations is linked to changes in both dynamics, network of contacts and energy profile in the protein-ligand complex. For the sulfonylurea (SU), the alterations in these attributes suggest a greater restoration of the competitive component of the inhibitor in the protein in relation to imidazolinone (IMI), in line with the greater inhibitory component itself reported in the literature for the SU compared to the IMI. In enzymes with the imidazolinone inhibitor, resistance seems to have a direct relationship with allosteric modifications, structurally modifying the cofactors region. In enzymes with the sulfonylurea inhibitor, the resistance pattern suggests a strong relationship with the loss of affinity for the ligand. The results obtained here can contribute to the elucidation of new paths for the sustainable theme of weeds, crops and herbicides. These same results also demonstrate that using machine learning to find patterns amidst a diversity of data returned by modeling and molecular dynamics is a readily applicable and effective strategy.

Keywords: acetolactatesynthases; ALS; herbicide resistance; machine learning; attribute selection.

Lista de Figuras

Figura 1. Etapas catalíticas para formação de isoleucina, valina e leucina. Fonte: Adaptado de [19].	21
Figura 2. (A) Processo de catálise realizada pelo cofator ThDP. (B) Monômero (Subunidade catalítica) de ALS colorido por seus domínios. Domínio α em ciano, β em magenta, γ em vermelho e região C-terminal em vermelho claro. Fonte: (A) adaptado de [19], (B) próprio autor.	22
Figura 3. Estrutura de homodímero de ALS. Herbicidas em esferas nos dois sítios ativos da interface de dimerização. Fonte: próprio autor.	23
Figura 4. Estrutura geral das sulfoniluréias. Fonte: adaptado de [36].	25
Figura 5. Estrutura geral das imidazolinonas. Fonte: adaptado de [13].	25
Figura 6. Comparação entre as técnicas de MTD e MD. O sistema é representado pelo círculo azul. Em uma MD o sistema permanece por muito tempo dentro de um mesmo FES enquanto que na MTD potenciais gaussianos (em rosa) são adicionados para estimular o sistema a mover-se para outro FES. Fonte: https://www.youtube.com/watch?v=CtIrLkx6aNo	29
Figura 7. Teoria do modo normal: (A) Descrição de um oscilador harmônico simples. Uma partícula m está ligada a uma mola com uma constante de força k e seu deslocamento x . (B) Análise de modo normal: aproximação harmônica da superfície de energia potencial. Para qualquer sistema biológico, a superfície de energia real é acidentada (linha pontilhada), mas para a análise de modo normal, a superfície é aproximada como uma superfície harmônica (linha simples). (C) Transformação de coordenadas cartesianas em coordenadas de modo normal. Neste exemplo de dois sistemas de coordenadas, as coordenadas do modo normal q são independentes, mas não as coordenadas cartesianas X . (D) Vetor de modo normal simples para a molécula de água. Cada seta representa a direção do movimento que cada átomo sofrerá conforme obtido da teoria do modo normal. O modo normal prevê três movimentos distintos para a molécula de água: modo simétrico, assimétrico e modo de flexão. Fonte: adaptado de [57]	30
Figura 8. RMSD ao longo da trajetória de cada simulação de MD relacionada ao seu respectivo frame inicial. Fonte: próprio autor.	42
Figura 9. RMSF ao longo da trajetória de MD, gráficos de sistemas mutantes não apresentados na figura. Cadeia A e B separadas no eixo x por barra preta e cinza, respectivamente. Fonte: próprio autor.	42

Figura 10. Diagrama de Venn entre os resíduos descritores de IMI e de SU. Não existem o compartilhamento dos resíduos descritores entre IMI e SU. Fonte: próprio autor.....	44
Figura 11. Localização dos resíduos descritores de IMI e SU na estrutura tridimensional da proteína. Monômeros coloridos de cores diferentes. Resíduos descritores encontrados na seleção de atributos em esferas verdes. Herbicidas em esferas cinza.	44
Figura 12. Rede alostérica e localização dos resíduos descritores de IMI e SU na estrutura tridimensional da proteína. Resíduos descritores em esferas verdes. Arestas coloridas (uma cor para cada posição de mutação) representam ligação entre a posição de mutação (saída das arestas) e os resíduos sensores/efetores dessa posição (direção das arestas). Fonte: próprio autor.	45
Figura 13. Matrizes de sobreposição entre os trinta primeiros PCs da trajetória WT/APO e trinta primeiros modos normais do primeiro frame da trajetória WT-APO. Fonte: próprio autor.	46
Figura 14. Gráfico de barras para valores de Fuga diagonal da análise de sobreposição entre os trinta primeiros PCs e trinta modos normais para sistema WT-APO. Fonte: próprio autor....	47
Figura 15. Autovalor das análises de PCA para o sistema WT/APO considerando os CA dos 24 descritores de IMI e dos 35 de SU. Fonte: próprio autor.	47
Figura 16. Autovalor da análise de modos normais para o sistema WT/APO considerando todos os CA. Fonte: próprio autor.....	48
Figura 17. Contribuição de cada PC das análises de PCA para o sistema WT/APO considerando os CA dos 24 descritores de IMI e dos 35 de SU. Fonte: próprio autor.....	48
Figura 18. Gráfico de barras para valores de sobreposição cumulativa entre os três primeiros PCs e modos normais para sistema WT-APO. Fonte: próprio autor.....	49
Figura 19. Gráficos de barras para valores de fuga diagonal calculados com resíduos descritores para todos os sistemas. Fonte: próprio autor.	50
Figura 20. Perfis de energia livre para cada sistema ligado a IMI e para sistema APO. Os triângulos ao longo dos gráficos representam as projeções médias de cada sistema nas PCs, sendo amarelos - resistentes, preto - holo, branco - apo e magenta - susceptíveis. Fonte: próprio autor.....	52
Figura 21. Perfis de energia livre para cada sistema ligado a SU e para sistema APO. Os triângulos ao longo dos gráficos representam as projeções médias de cada sistema ao longo das PCs, sendo amarelos - resistentes, preto - holo, branco - apo e magenta - susceptíveis. Fonte: próprio autor.	53

Figura 22. Perfis de energia livre para o conjunto de trajetórias IMI, levando em consideração os CA de todos os resíduos e os CA de resíduos descritores. Os triângulos ao longo do gráfico representam as projeções médias de cada sistema e cluster mais populoso de cada sistema ao longo das PCs. Triângulos amarelos - resistentes, preto - holo, branco - apo e magenta - susceptíveis..... 54

Figura 23. Perfis de energia livre para o conjunto de trajetórias SU, levando em consideração os CA de todos os resíduos e os CA de resíduos descritores. Os triângulos ao longo do gráfico representam as projeções médias de cada sistema e cluster mais populoso de cada sistema ao longo das PCs. Triângulos amarelos - resistentes, preto - holo, branco - apo e magenta - susceptíveis..... 55

Figura 24. Movimentos adotados por resíduos descritores de sistemas ligados a IMI nas componentes principais PC1, PC2, PC3. Monômeros coloridos de azul escuro e cinza escuro, esferas roxas transparentes representam um herbicida ligado ao sítio. Setas alaranjadas representam a direção do movimento previsto para os resíduos descritores. Fonte: próprio autor. 56

Figura 25. Movimentos adotados por resíduos descritores de sistemas ligados a SU nas componentes principais PC1, PC2, PC3. Monômeros coloridos de azul escuro e cinza escuro, esferas roxas transparentes representam um herbicida ligado ao sítio. Setas alaranjadas representam a direção do movimento previsto para os resíduos descritores. Fonte: próprio autor. 57

Figura 26. Gráfico de barras para valores de $\Delta\Delta G$ (ΔG do sistema holo/mutantes - ΔG de sistemas WT/holo) para cada sistema com IMI. Barras pretas - sistemas descritos na literatura como resistentes, barras cinza claro - sistemas descritos na literatura como susceptíveis, barras cinza escuro - sistemas não discriminados na literatura. * apontam para sistemas onde a técnica não foi assertiva. Triângulos azuis escuro - mutação de resistência para IMI e não SU, azul claro - mutação de resistência para IMI e não informado para SU, magenta - mutação de resistência para SU e não IMI, rosa claro - mutação de resistência para SU e não informado para IMI e verde - mutação de resistência para IMI e SU. Fonte: próprio autor..... 58

Figura 27. Gráfico de barras para valores de $\Delta\Delta G$ (ΔG do sistema holo/mutantes - ΔG de sistemas WT/holo) para cada sistema com SU. Barras pretas - sistemas descritos na literatura como resistentes, barras cinza claro - sistemas descritos na literatura como susceptíveis, barras cinza escuro - sistemas não discriminados na literatura. * apontam para sistemas onde a técnica não foi assertiva. Triângulos azuis escuro - mutação de resistência para IMI e não SU, azul claro - mutação de resistência para IMI e não informado para SU, magenta - mutação de resistência

para SU e não IMI, rosa claro - mutação de resistência para SU e não informado para IMI e verde - mutação de resistência para IMI e SU. Fonte: próprio autor..... 59

Figura 28. Gráfico de barras para valores de (A) $\Delta\Delta G$ do complexo ternário, (B) $\Delta(\Delta G$ de cooperatividade), (C) $\Delta\Delta G$ de associação do segundo ligante e (D) $\Delta\Delta G$ da metadinâmica, para sistemas ligados a IMI. Barras pretas - sistemas descritos na literatura como resistentes, barras cinza-claro - sistemas descritos na literatura como susceptíveis, barras cinza escuro - sistemas não discriminados na literatura. * apontam para sistemas onde a técnica não foi assertiva. Triângulos azuis escuro - mutação de resistência para IMI e não SU, azul claro - mutação de resistência para IMI e não informado para SU, magenta - mutação de resistência para SU e não IMI, rosa claro - mutação de resistência para SU e não informado para IMI e verde - mutação de resistência para IMI e SU. Círculos verdes $\Delta\Delta G$ não cooperativo, círculos laranja $\Delta\Delta G$ antiooperativo e círculos amarelos $\Delta\Delta G$ cooperativo. Fonte: próprio autor. 60

Figura 29. Gráfico de barras para valores de (A) $\Delta\Delta G$ do complexo ternário, (B) $\Delta(\Delta G$ de cooperatividade), (C) $\Delta\Delta G$ de associação do segundo ligante e (D) $\Delta\Delta G$ da metadinâmica, para sistemas ligados a SU. Barras pretas - sistemas descritos na literatura como resistentes, barras cinza-claro - sistemas descritos na literatura como susceptíveis, barras cinza escuro - sistemas não discriminados na literatura. * apontam para sistemas onde a técnica não foi assertiva. Triângulos azuis escuro - mutação de resistência para IMI e não SU, azul claro - mutação de resistência para IMI e não informado para SU, magenta - mutação de resistência para SU e não IMI, rosa claro - mutação de resistência para SU e não informado para IMI e verde - mutação de resistência para IMI e SU. Círculos verdes $\Delta\Delta G$ não cooperativo, círculos laranja $\Delta\Delta G$ antiooperativo e círculos amarelos $\Delta\Delta G$ cooperativo. Fonte: próprio autor. 62

Figura 30. Contatos retornados após a seleção de atributos para IMI na estrutura da proteína WT/holo. Carbonos de IMI em magenta, de FAD em roxo, de ThDP em salmão e de resíduos da proteína em verde. Nitrogênio em azul, hidrogênios em branco, oxigênios em vermelho, fosforo em laranja e enxofre em amarelo. Linhas pontilhadas representam a distância entre os atômicos que fazem contato. Fonte: próprio autor. 65

Figura 31. Contatos retornados para SU entre proteína-FAD e proteína_THDP na estrutura da proteína WT/holo e holo/mutantes resistente. Carbonos de SU em magenta, de FAD em roxo, de ThDP em salmão e de resíduos da proteína em verde. Nitrogênio em azul, hidrogênios em branco, oxigênios em vermelho, fosforo em laranja e enxofre em amarelo. Linhas pontilhadas representam a distância entre os atômicos que fazem contato. Distância 4.6 nos contatos de WT/holo e de 3.6 e 3.9 nos contatos de holo/mutante/resistente. Fonte: próprio autor. 66

Figura 32. Contatos retornados para SU na estrutura da proteína WT/holo e holo/mutantes resistente. Carbonos de SU em magenta, de FAD em roxo, de ThDP em salmão e de resíduos da proteína em verde. Nitrogênio em azul, hidrogênios em branco, oxigênios em vermelho, fosforo em laranja e enxofre em amarelo. Linhas pontilhadas representam a distância entre os atômicos que fazem contato. Distância 4.6 nos contatos de WT/holo e de 3.6 e 3.9 nos contatos de holo/mutante/resistente. Fonte: próprio autor.67

Lista de Tabelas

Tabela 1. Mutações inseridas nos modelos com herbicidas. Numeração dos resíduos de acordo com ALS de <i>Arabidopsis thaliana</i>	34
Tabela 2. Resultados obtidos para o classificador MLP de IMI e de SU.	43
Tabela 3. Contatos importantes para a classificação em resistentes e susceptíveis retornados pela seleção de atributos para sistemas com IMI.	63
Tabela 4. Contatos importantes para a classificação em resistentes e susceptíveis retornados pela seleção de atributos para sistemas com SU.....	64
Tabela 5. Descritores gerais retornados pela seleção de atributos.....	68
Tabela 6. Resultados obtidos para o classificador MLP da junção dos melhores contatos com melhores descritores gerais de IMI.....	69
Tabela 7. Resultados obtidos para o classificador MLP da junção dos melhores contatos com melhores descritores gerais de SU.....	69

Lista de Abreviações

AHAS: Acetohidroxiácido sintase

ALS: Acetolactato Sintase

Eho: *Vibrational energy*

FAD: Dinucleotídeo de flavinaadenina

GAFF: *General Amber Force Field*

HEThDP: Hidroxietilamina

IMI: Imidazolinonas

MD: *Molecular dynamics*

MDT: *Metadynamics*

MM/PBSA: *Molecular Mechanics Poisson-Boltzmann Surface Area*

MLP: *Multilayer perceptron*

MOA: *Mechanisms of action*

PCA: *Principal component analysis*

PDB: *Protein Data Bank*

PME: *Particle Mesh Ewald*

PRS: *Perturbation Response Scanning*

PTB: Pirimidiniltiobenzoatos

RMSF: *Root mean square fluctuation*

RMSD: *Root meansquare deviation*

SCT: Sulfonilaminocarboniltriazolinona

Sho: *Vibrational entropy*

SU: Sulfonilureias

TDM: Tiamina no estado intermediário

ThDP: Difosfato de Tiamina

TP: Triazolopirimidinasulfonanilidas

WT: Wild Type

ΔG : *Gibbs freeenergy*

Sumário

1. Introdução.....	18
1.1 Plantas daninhas e herbicidas	18
1.2 Resistência a herbicidas.....	19
1.3 ALSs.....	20
1.4 ALS e resistência a herbicidas.....	23
1.5 Sulfoniluréia e Imidazolinonas.....	24
1.6 Bioinformática Estrutural	26
1.7 Simulações de dinâmica molecular de equilíbrio e de não-equilíbrio (o exemplo da técnica de metadinâmica)	26
1.8 Análise de modos normais.....	29
1.9 Aprendizado de Máquina.....	30
2. Objetivos.....	32
2.1 Objetivo geral	32
2.2 Objetivos específicos	32
3. Materiais e métodos.....	33
3.1 Coleta de dados/Modelagem dos sistemas	33
3.2 Simulações de MD e Simulações de metadinâmica com escape do ligante.....	34
3.3 Predição de resíduos descritores por aprendizado de máquina	36
3.4 Redes Alostéricas	37
3.5 Análise de modos normais de Componente principal (PCA).....	37
3.6 Análise de afinidade pelo ligante.....	38
3.7 Reconstrução dos perfis de energia livre e PCA	40
3.8 Análise Farmacofórica.....	40
3.9 Aprendizado de máquina e dados de MD.....	41

4. Resultados e discussões	42
4.1 Análises de RMSD e RMSF das trajetórias de MD	42
4.2 Aprendizado de máquina e resíduos descritores de mobilidade.....	43
4.3 Redes alostéricas.....	44
4.4 Análise de sobreposição de PCA e modos normais	45
4.5 Análise de PCA	51
4.5.1 PyReweighting	51
4.5.2 Movimentos das componentes principais.....	56
4.6 Análises de afinidade pelo ligante	57
4.6.1 Análise de metadinâmica.....	57
4.6.2 MM/PBSA e Quase-harmônica.....	59
4.7 Análise farmacofórica.....	62
4.8 Aprendizado de máquina e dados de MD.....	67
5. Conclusões.....	70
6. Perspectivas	72
7. Referências bibliográficas	73

1. Introdução

1.1 Plantas daninhas e herbicidas

Desde os primórdios da agricultura, o ser humano como produtor vem lidando com situações adversas, passando a enfrentar não só fatores abióticos, como o clima, falta ou excesso de água, como também fatores bióticos, como insetos, pássaros e as ervas daninhas [1]. Atualmente a agricultura mundial provê alimento para mais de 7 bilhões de pessoas [2].

Campos de cultivo agrícola são ambientes onde se mantém um agroecossistema bem definido, com suas particularidades de um ambiente com distúrbios intensos e extensos e estresses inseridos pelo homem. Esses campos são o berço para o crescimento de plantas adaptativas, conhecidas como plantas invasoras e ou ervas daninhas. Elas utilizam esse ambiente que aumenta sua capacidade reprodutiva, denominada estratégia adaptativa [3]. A presença de plantas invasoras prejudica não só o rendimento da safra, como também a qualidade da mesma. Estas competem por nutrientes, luz solar entre outros elementos essenciais [4].

Medidas de proteção de safras foram então adotadas e se fazem necessárias na pré-colheita (aplicação em campo) e na pós-colheita (aplicação no armazenamento) [1]. As medidas de proteção em campo devem receber maior atenção, podendo as perdas de safra nessa fase ser mais agressivas. Medidas de proteção são adotadas constantemente entre os produtores, sendo o mais comum o uso de pesticidas [2].

Aproximadamente 60% dos pesticidas utilizados globalmente são herbicidas destinados à eliminação das ervas daninhas consideradas como uma ameaça bem difundida à segurança alimentar e à sustentabilidade agrícola. Aproximadamente 34% das perdas de rendimento estão relacionadas a ervas daninhas [1,5,6].

Em 1900 introduziu-se o uso sais de cobre inorgânico como herbicida. Já no ano de 1940, surgem os herbicidas sintéticos, que até a atualidade continuam sendo a principal estratégia de controle contra ervas daninhas. Herbicidas sintéticos são, portanto, pequenas moléculas desenvolvidas para atingirem um alvo ou um processo específico de uma planta a fim de interromper seu crescimento [1,3,5].

Após o desenvolvimento de herbicidas sintéticos, as próximas décadas (1950 a meados de 1980) foram marcadas pelo lançamento de novos químicos com novos mecanismos de ação (*mechanisms of action*- MOA) a cada 2 anos. No entanto, a partir da década de 80 nenhum novo MOA foi registrado [5,6].

Os herbicidas podem ser subdivididos em três principais grupos com base em seus MOA: (i) direcionados para vias bioquímicas e processos fisiológicos da fotossíntese; (ii) direcionados para inibir a formação de açúcares, aminoácidos, ácidos graxos, ou seja, blocos de construção biológicos; e (iii) outros mecanismos de ação (como hormonal e síntese proteica). Para que um herbicida seja eficaz, ele deve ser capaz de atingir seu alvo em dose letal, interrompendo assim parte importante do metabolismo da planta [2,5].

1.2 Resistência a herbicidas

Dificuldades enfrentadas durante o desenvolvimento e produção de novos químicos, como por exemplo o tempo (aproximadamente 10 anos) e o custo (aproximadamente 300 milhões de dólares), levaram à aplicação de um mesmo químico constantemente às culturas. O fato de que as culturas de interesse são preparadas, ou seja, geneticamente modificadas a fim de não serem atingidas pelos herbicidas (características de resistência), também favorece o uso de químicos específicos por longos períodos de tempo [7,8,9].

Esses herbicidas sintéticos aplicados às culturas levaram à evolução de plantas daninhas de biótipos resistentes a herbicidas. A pressão de seleção foi fortemente experimentada pelas ervas daninhas. Biótipos resistentes possuem mutações em seu DNA que o permitem sobreviver e passar para sua prole a característica de resistência, sendo então selecionados em populações existentes em meio a culturas sob aplicação de herbicidas. Esses biótipos apresentam geralmente resistência local-alvo, sendo documentadas mutações pontuais em várias espécies [8,9,10]. Essas mutações são geralmente encontradas na sequência de enzimas chave para a inibição pelos herbicidas [11,12]. Em particular, é bem documentado que substituições do tipo único aminoácido podem ser significativamente eficazes na alteração da responsividade de proteínas a diferentes ligantes e estímulos, sem no entanto prejudicar o enovelamento e funcionalidade globais [81].

De acordo com a Pesquisa Internacional de ervas daninhas resistentes a herbicidas, atualmente existem 26 locais de ação de herbicidas conhecidos e as plantas daninhas desenvolveram resistência a pelo menos 23 desses locais e a 167 herbicidas diferentes. Atualmente são relatados 514 casos únicos de ervas daninhas resistentes (espécie versus local de ação) [13]. Dentre os 514 casos únicos atualmente relatados de ervas daninhas resistentes, 165 destes foram a herbicidas inibidores da enzima Acetolactato Sintase (ALS), sendo o maior grupo de resistência [13].

1.3 ALSs

As enzimas Acetolactato sintases (ALSs), também conhecidas como acetohidroxiácidos sintases (AHAS; EC 2.2.1.6), são enzimas encontradas em todas as plantas, bactérias e fungos. Estão envolvidas na via de biossíntese de aminoácidos de cadeia ramificada (valina, leucina e isoleucina), sendo a primeira enzima na via (**Figura 1**). Nas reações catalizadas por esta enzima duas moléculas de piruvato são conjugadas e convertidas a 2-acetolactato (via de síntese para valina e leucina) ou uma molécula de piruvato e uma de 2-cetobutirato são convertidas a 2-aceto-2-hidroxiacetato (via para isoleucina) [14,15].

As ALSs são formadas por duas subunidades, uma subunidade catalítica e outra subunidade reguladora. A subunidade catalítica possui massa molecular de aproximadamente 60-70 kDa e a subunidade reguladora entre 9 e 54 kDa dependendo da espécie. A subunidade catalítica só é completamente funcional quando se une a uma segunda subunidade catalítica (idêntica) formando um homodímero para a realização da catálise [16,17].

A subunidade catalítica faz parte da superfamília de enzimas dependentes de difosfato de Tiamina (ThDP), forma bioativa da vitamina B1, um cofator que é essencial para a catálise realizada pela enzima. O ThDP ataca uma molécula de piruvato que é descarboxilado, agora como o intermediário hidroxietilamina (HETHDP) enamina/carbânion se condensa a outra molécula de piruvato para formar 2-acetolactato ou a uma molécula 2-cetobutirato formando 2-aceto-2-hidroxiacetato (**Figura 2(A)**) [18,19,20,21].

As ALSs ainda necessitam de mais dois cofatores, um íon magnésio (Mg^{2+}) para coordenar (ancorar) o ThDP e um dinucleotídeo de flavina adenina (FAD), sendo que este último não se envolve na catálise e é tratado como um resquício da família de enzimas dependentes de ThDP (**Figura 2(B)**) [14,22,23].

A enzima em sua forma ativa se encontra nos cloroplastos de organismos vegetais. É composta por 4 subunidades idênticas, formando um tetrâmero. Cada tetrâmero possui 4 sítios catalíticos, sendo que o sítio é marcado pela presença do cofator ThDP em cada subunidade. Cada unidade homodimérica possui dois sítios catalíticos idênticos em faces opostas na interface da homodimerização. Sendo assim, um monômero completa o sítio catalítico do outro para se tornar ativo. A figura **Figura 3** mostra um herbicida em cada um dos monômeros na interface da dimerização [19].

Cada monômero é composto por 3 domínios. Domínio α (resíduos 87–280), β (resíduos 281–451), γ (resíduos 463–639) e região C-terminal (640–668). Os primeiros 86 resíduos fazem parte de um peptídeo de trânsito para enviar cada subunidade ao cloroplasto. (figura do dímero) [19].

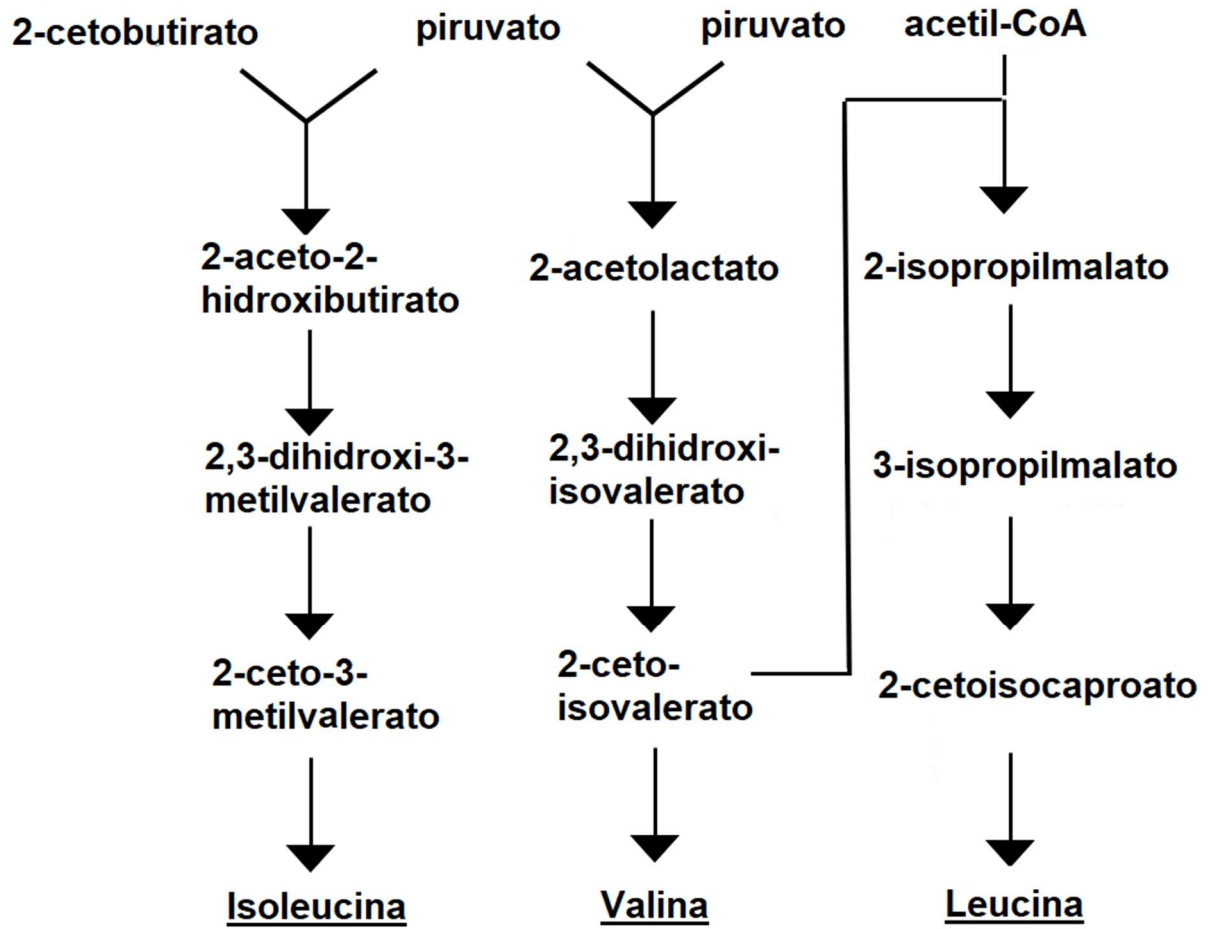


Figura 1. Etapas catalíticas para formação de isoleucina, valina e leucina.

Fonte: Adaptado de [19].

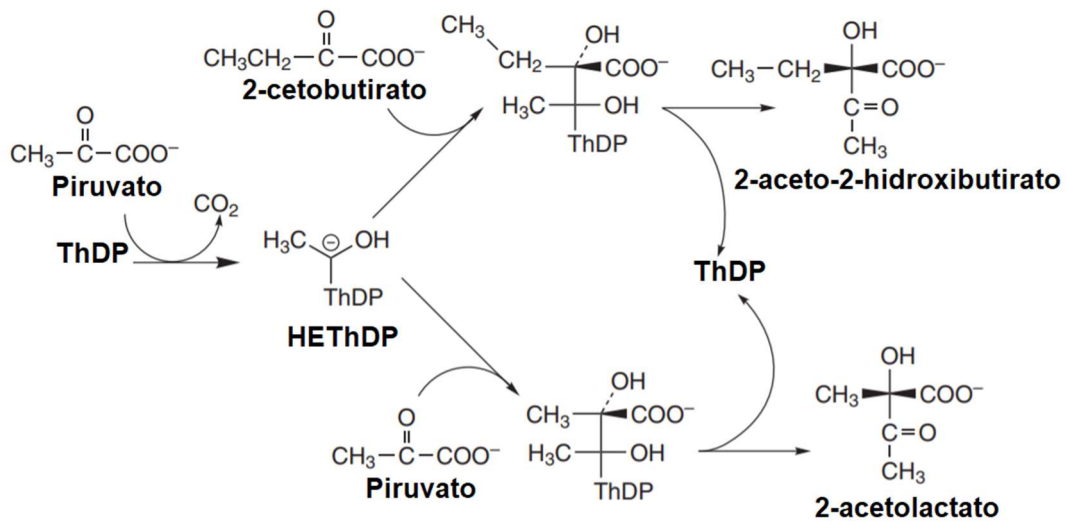
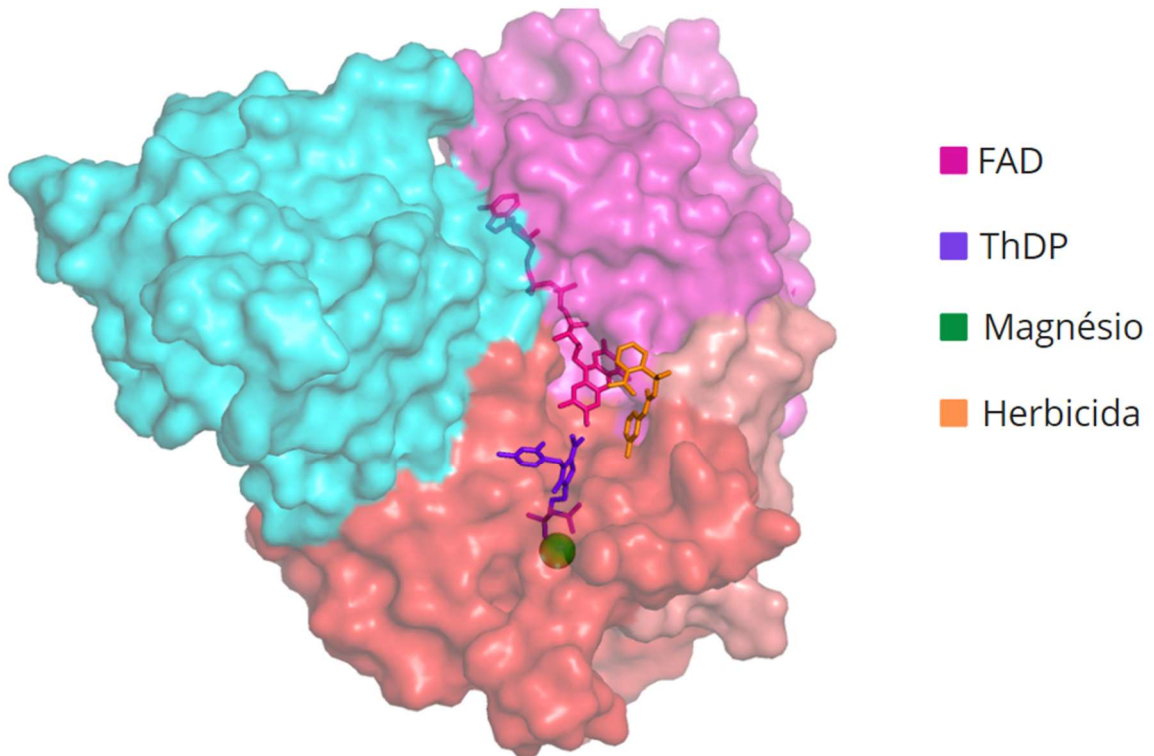
A**B**

Figura 2. (A) Processo de catálise realizada pelo cofator ThDP. **(B)** Monômero (Subunidade catalítica) de ALS colorido por seus domínios. Domínio α em ciano, β em magenta, γ em vermelho e região C-terminal em vermelho claro.

Fonte: (A) adaptado de [19], (B) próprio autor.

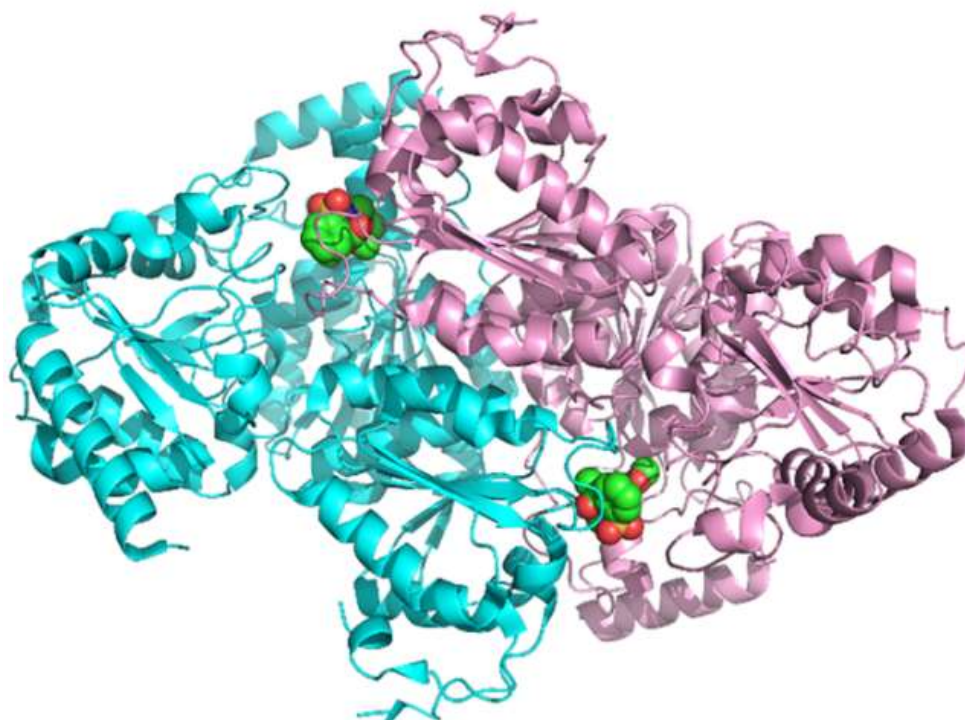


Figura 3. Estrutura de homodímero de ALS. Herbicidas em esferas nos dois sítios ativos da interface de dimerização.

Fonte: próprio autor.

1.4 ALS e resistência a herbicidas

O desenvolvimento e introdução da primeira classe de herbicidas (classe das sulfonilureias) direcionada a enzimas ALSs marcou o histórico da problemática com ervas daninhas. A aplicação desses químicos desde sua inserção em 1982 até a atualidade é realizada em baixas doses de ingrediente ativo ($\sim 10\text{--}100\text{ g}\cdot\text{ha}^{-1}$), possuindo também baixa toxicidade em mamíferos e ampla seletividade de culturas [24,25].

Atualmente são comercializadas 5 classes de herbicidas inibidores de ALS:

1. Sulfonilureias (SU), o primeiro químico a ser comercializado,
2. Imidazolinonas (IMIs),
3. Triazolopirimidina sulfonanilidas (TPs)
4. Pirimidiniltiobenzatos (PTBs) e
5. Sulfonilaminocarboniltriazolinona (SCT) [26].

O primeiro caso de resistência local alvo a herbicidas inibidores de ALS foi relatado nos Estados Unidos no ano de 1987 em *Lactuca serriola* L., apenas 5 anos após sua inserção no mercado. Desde então inúmeros casos de resistência foram relatados para as 5 diferentes classes [27].

Como mencionado no item 1.2, várias mutações pontuais acarretam em uma resistência local-alvo, sendo descritos na literatura resíduos de aminoácidos que, ao serem mutados, permitem resistência a herbicidas.

A sequência da enzima ALS se mostra conservada nos diferentes tipos de filós onde é encontrada, e em particular em organismos vegetais. A enzima possui identidade geral superior a 63% e nos casos de plantas uma média de 80%. Cinco resíduos geralmente conservados na sequência de ALS de organismos vegetais são alvo de mutações que garantem resistência a herbicidas. Seguindo a numeração de aminoácidos de *Arabidopsis thaliana* (L.) Heynh. fornecida por Tranel e Wright [24], três posições de mutação (das cinco conservadas) estão localizadas no domínio N-terminal: ALA122, PRO197, ALA205 e as outras duas na extremidade terminal carboxil: TRP574 e SER653. Alguns outros resíduos que podem ser mutados e garantir resistência estão localizados nas proximidades destes cinco mais conservados; são eles: ASP376, ARG377, LYS434, GLU268 [24,27,28].

Os resíduos alvo de mutação podem ser substituídos por vários outros resíduos para conferir resistência. Substituições de PRO197 foram relatadas para doze resíduos, ALA122 para quatro outros resíduos, e SER653 para três outros [25,29,30].

Dentre as 5 classes de inibidores de ALS, as sulfoniluréias e imidazolinonas são mais amplamente comercializadas em todo o mundo. Não são tóxicos aos animais e devido a sua seletividade e eficácia, podem ser aplicados em doses baixas [23].

Algumas mutações possuem a característica de resistência cruzada a mais de um herbicida. Mutação W574L e D376E conferem resistência a moléculas de imidazolinonas e a sulfonilureias. Já mutações na PRO197 conferem resistência a apenas Sulfonilureia e SER653 e ALA122 apenas a Imidazolinonas [25,27].

Neste contexto, surge a necessidade de se compreender a natureza da resistência aos diferentes tipos de herbicidas atualmente comercializados, podendo-se assim iniciar um tratamento mais assertivo contra as ervas daninhas. Técnicas computacionais podem ser uma opção mais rápida e econômica para compreensão das particularidades que envolvem a resistência a herbicidas, sendo a bioinformática estrutural uma peça chave para esse tipo de abordagem.

1.5 Sulfoniluréia e Imidazolinonas

Dentro do grupo das sulfoniluréias são relatados mais de 30 compostos ativos atualmente comercializados [13]. O Termo "sulfoniluréia" é derivado da porção central da molécula, ou ponte, que é composta de sulfonilureia ligada a um anel aromático pelo átomo de

enxofre, e um anel heterocíclico pelo de nitrogênio. Uma ampla gama de substituições é possível dentro das sulfonilureias, visualizadas nos dois sistemas de anel (**Figura 4**)[45,46].

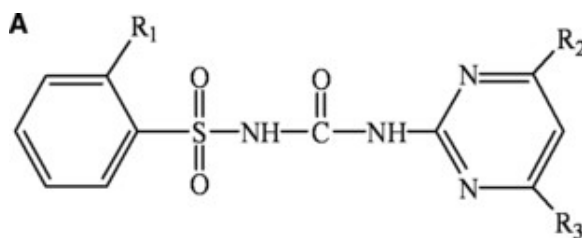


Figura 4. Estrutura geral das sulfoniluréis.

Fonte: adaptado de [36].

Dentro do grupo das imidazolinonas encontram-se seis compostos ativos e comercializados atualmente [13]. A maioria das imidazolinonas comerciais consiste em um anel de piridina carboxilado, que pode ser substituído em R1 e/ou R2 (**Figura 5**). O anel de piridina também pode ser substituído (X). A melhor atividade herbicida é obtida com um anel dihidroimidazolona que é substituído por grupos metil e isopropil.

Ambos os herbicidas possuem a característica de inibição reversível caracterizada pela ligação do inibidor ao sítio ativo ou a outra parte da cadeia polipeptídica da enzima, impedindo que o substrato seja transformado em produto. Além de serem inibidores reversíveis, os herbicidas IMI e SU foram relatados como sendo mistos e um pouco mais competitivos no caso de SU [50,52,53,95,96,97].

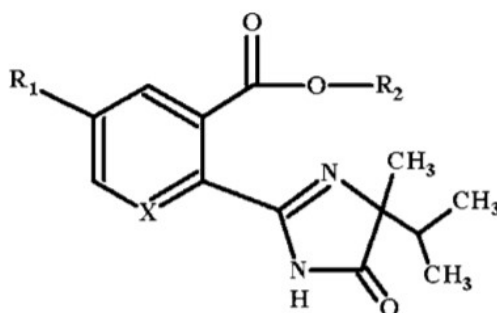


Figura 5. Estrutura geral das imidazolinonas.

Fonte: adaptado de [13].

Inibidores mistos possuem características de inibidores competitivos e não competitivos. Inibidores competitivos competem pelo sítio ativo, interferem na afinidade do substrato ao sítio ativo. Quando o inibidor se liga ao sítio, a enzima não é capaz de receber o substrato o que reflete em um aumento de Km. Inibidores incompetitivos são capazes de modificar estruturalmente o sítio ativo de modo que, o substrato consegue acessar o sítio, mas

a reação catalítica ocorre de maneira mais lenta, refletindo na diminuição dos valores de K_{cat} [54].

1.6 Bioinformática Estrutural

Bioinformática estrutural é uma subdisciplina da Bioinformática com o intuito de compreender melhor a função de um biopolímero através de sua estrutura tridimensional e não apenas de sua sequência. Sendo assim, informações contidas em uma estrutura tridimensional podem ser organizadas e classificadas de acordo com suas capacidades funcionais [31].

Um marco no histórico de desenvolvimento da bioinformática estrutural se deu em 1971 com o estabelecimento do *Protein data bank* - PDB (www.rcsb.org/), o primeiro recurso de dados moleculares de acesso aberto em biologia, servindo como um repositório global de estruturas tridimensionais a nível atômico. No ano de 1976 o PDB continha 23 estruturas. Atualmente mais de 170.000 estruturas determinadas experimentalmente, entre proteínas, fragmentos de DNA e RNA estão disponíveis no acervo [32,33,34].

A rápida evolução de técnicas cada vez mais acuradas para a determinação da estrutura de proteínas (sua maioria por difração de raios-x) e também o acesso rápido, fácil e gratuito a arquivos de coordenadas das mesmas, vem sendo um diferencial para trabalhos computacionais. Várias técnicas vêm sendo descobertas e aperfeiçoadas desde então [31].

Análises envolvendo a manipulação das coordenadas estruturais, principalmente de proteínas, podem trazer vários *insights* sobre a função e mecanismos de ligação com outras moléculas. Neste trabalho algumas análises desse porte foram realizadas e então se faz necessário um maior detalhamento das mesmas nas próximas sessões.

1.7 Simulações de dinâmica molecular de equilíbrio e de não-equilíbrio (o exemplo da técnica de metadinâmica)

Vários estudos mostraram que as moléculas de proteína não são corpos rígidos, mas sim flexíveis. Movimentos específicos podem ser cruciais para o seu funcionamento. A flexibilidade das proteínas tem sido associada a vários aspectos funcionais, como ativação da atividade catalítica, transdução de sinal e mecanismos alostéricos [31,58].

Simulações de dinâmica molecular (MD, do inglês *Molecular Dynamics*) clássicas (isto é, que não utilizam métodos quânticos) têm o propósito de descrever um dado sistema molecular de alta complexidade ao longo do tempo, tratando seus potenciais e dinâmica por aproximações da mecânica newtoniana. Esta aproximação é, via de regras, prontamente factível para sistemas basicamente orgânicos e em meio aquoso nas Condições Normais de Temperatura e Pressão (CNTP) e para os quais o interesse maior seja para processos não diretamente

dependentes de reações químicas [82,83]. Vários trabalhos vêm relatando a utilização de simulações de MD para compreensão de processos envolvendo enovelamento de proteínas, estabilidade, interação proteína-proteína e também interação proteína-ligante [59,60,61,62].

Nas MD clássicas (doravante denotadas apenas como MD), calcula-se a força que atua sobre cada átomo a partir da derivada da função energia potencial. Os átomos de um biopolímero se movem de acordo com as equações de movimento de Newton.

$$m_{\alpha} \ddot{\vec{r}}_{\alpha} = - \frac{\partial}{\partial \vec{r}_{\alpha}} U_{\text{total}}(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N), \quad \alpha = 1, 2 \dots N,$$

Equação 1

Assim, a simulação propaga a posição r_{α} de um átomo α no sistema de interesse contendo N partículas ao atribuir a este átomo uma aceleração \ddot{r}_{α} dependente de sua massa m_{α} , e da força obtida como derivada da energia potencial total U_{total} atuante sobre o mesmo átomo. Tal potencial U_{total} , por sua vez, é dependente do conjunto de todas as posições atômicas no mesmo sistema em um certo instante t . Além disto, o mesmo potencial é decomposto em uma série de aproximações harmônicas e cossenoides simples envolvendo conjuntos de átomos covalentemente ligados, bem como de Lennard-Jones (interações de Van der Waals a curta distância) e de Coulomb (interações eletrostáticas a longa distância) para átomos separados por mais de quatro ligações covalentes (sendo tais aproximações clássicas dos potenciais intra e intermoleculares sendo denotadas como *campos de força*) (**Equação 2**) [63,83].

$$U = \sum_{\text{lig}} k_b (r - r_0)^2 + \sum_{\text{ang}} k_a (\theta - \theta_0)^2 + \sum_{\text{died}} [1 + \cos(n\phi - \delta)] \\ + \sum_{\text{L.J}} 4\epsilon_{ij} \left(\frac{\sigma_{ij}^{12}}{r_{ij}^{12}} - \frac{\sigma_{ij}^6}{r_{ij}^6} \right) + \sum_{\text{elet}} \frac{q_i q_j}{r_{ij}}$$

Equação 2

Portanto as simulações por MD requerem a execução de um grande número de tarefas, como o cálculo de forças dos átomos que estão quimicamente ligados e dos átomos que não estão ligados covalentemente assim como ângulos e torção; atribuição de velocidades iniciais aleatórias para cada átomo do sistema (respeitando-se, contudo, uma distribuição de Boltzmann compatível com a temperatura adotada); atualização da nova posição do átomo, cálculo da energia potencial e da força total incidida sobre cada átomo, atualização das velocidades e

posições de cada átomo ou molécula do sistema sendo simulado. Essas tarefas são repetidas várias vezes por um tempo pré-determinado pelo usuário [63].

Alguns processos biológicos, como a ligação de uma proteína ao seu ligante, são eventos de difícil visualização e de difícil amostragem com técnicas padrão de MD em equilíbrio, onde processos de longa escala são improváveis de se ver durante a simulação. Estudos de estimativa de afinidade de ligação ao ligante podem então ser melhor direcionados através de novas técnicas emergentes de não-equilíbrio (i.e., com a adição de algum potencial externo definido pelo usuário) como a Metadinâmica (MTD) [64].

Em uma MD, quando o sistema alcança uma configuração cercada por barreiras energéticas significativas ele tende a ficar “aprisionado” dentro deste estado *metaestável*. Para simulações de MD em intervalos de tempo mais praticáveis (na faixa dos nanossegundos), o sistema tende a ficar preso nesse “vale” de equilíbrio ao longo da superfície de energia livre (FES, do inglês Free Energy Surface) e não amostrar outros estados de igual, ou de maior importância [84,85]. Nesse contexto a MTD adiciona um potencial tendencioso para empurrar o sistema para fora desse vale e então amostrar outros estados (**Figura 6**) [64].

A essência da MTD está em adicionar periodicamente, um potencial de polarização gaussiano repulsivo de altura e largura determinados pelo usuário à superfície de energia potencial. O usuário informa algumas variáveis coletivas, que são coordenadas que vão representar o processo ao longo de uma via de reação. O viés gaussiano está centrado nas posições (coordenadas) das variáveis coletivas que o sistema está visitando no momento. Esse viés vai preenchendo e se acumulando dentro dos “vales” de superfície de energia livre estimulando o sistema a transpor barreiras e visitar outros espaços [64]. A MTD foi ludicamente descrita como “encher os poços de energia livre com areia computacional” o que permite uma amostragem de eventos antes não amostrados pela MD.

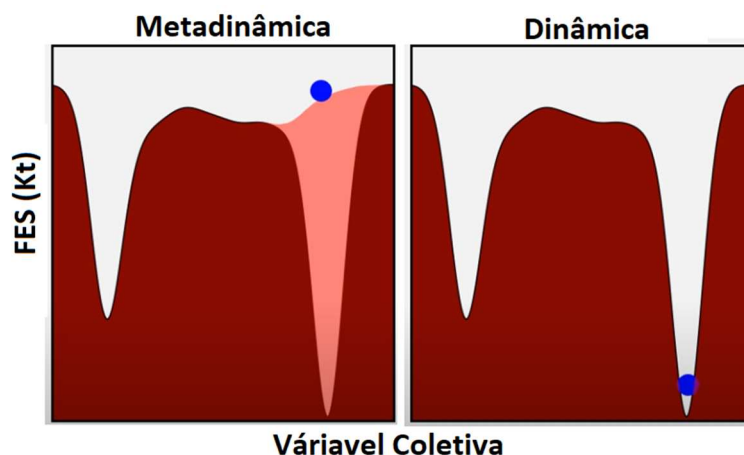


Figura 6. Comparação entre as técnicas de MTD e MD. O sistema é representado pelo círculo azul. Em uma MD o sistema permanece por muito tempo dentro de um mesmo FES enquanto que na MTD potenciais gaussianos (em rosa) são adicionados para estimular o sistema a mover-se para outro FES.

Fonte: <https://www.youtube.com/watch?v=CtIrLkx6aNo>.

1.8 Análise de modos normais

Análises de modos normais também é voltada para previsão e análise de movimentos coletivos em proteínas que em sua grande maioria estão envolvidos com a função ou interação com outras biomoléculas. Diferentemente da técnica de MD o gasto computacional é bem reduzido, sendo as equações de movimento resolvidas de forma analítica e não numérica.

A análise de modos normais, portanto, investiga o movimento vibracional de um sistema harmônico que é oscilante na vizinhança de seu ponto de equilíbrio. Esses movimentos são de pequena amplitude, permanecendo num poço de potencial energético não podendo ultrapassar barreiras energéticas. Quando o sistema está no fundo desse poço de potencial, ele se encontra em equilíbrio, ou seja, a força agindo no sistema é igual a 0 [55].

Os modos normais vibracionais nos fornecem o padrão dos movimentos, através dos autovetores e autovalores associados. Os vetores dos modos normais descrevem em que direção cada partícula do sistema se move, e o quanto ela se move em relação as outras partículas. A vibração de cada partícula do sistema é na mesma frequência em cada um dos modos calculados. Os modos vibracionais de baixa energia, correspondem a movimentos de caráter coletivo, enquanto os modos de alta frequência correspondem a distorções locais (**Figura 7**) [55,56,57].

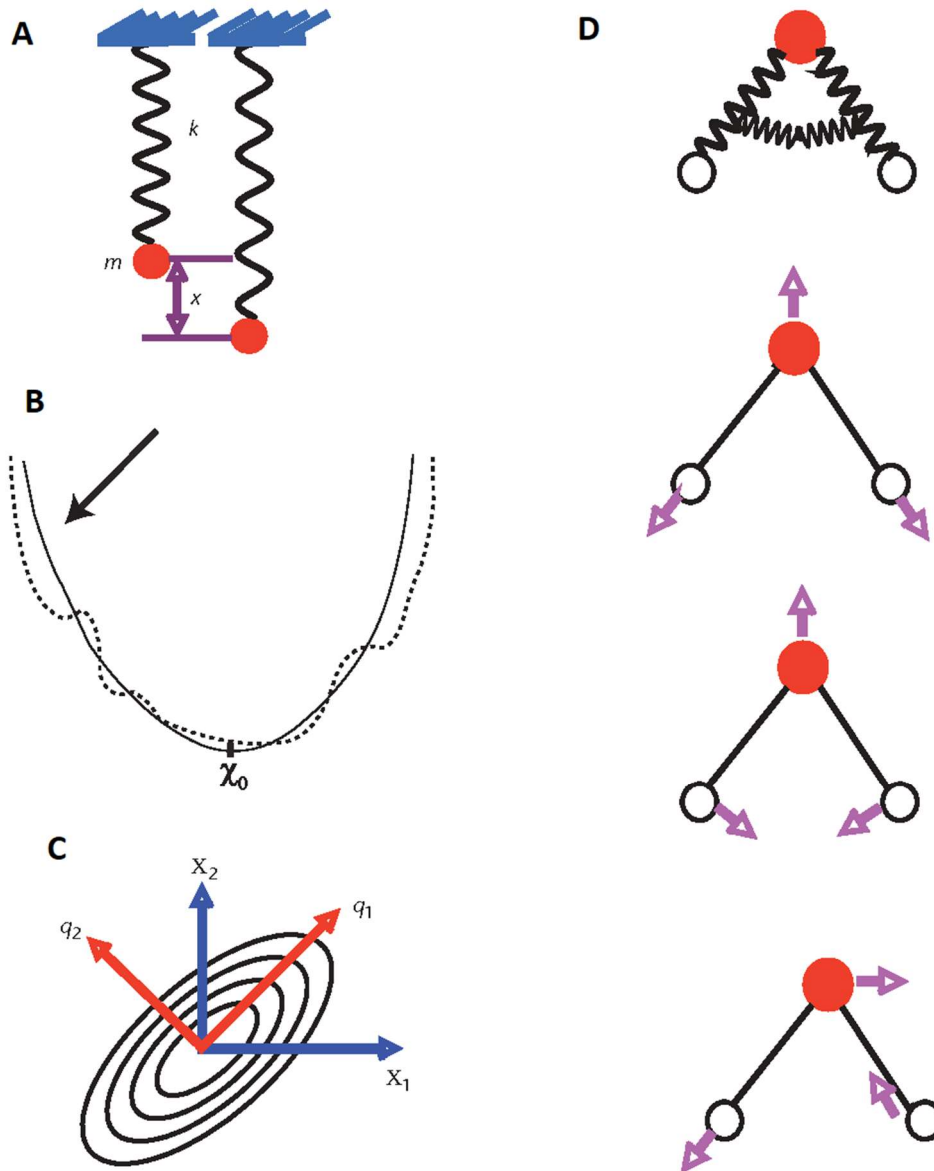


Figura 7. Teoria do modo normal: (A) Descrição de um oscilador harmônico simples. Uma partícula m está ligada a uma mola com uma constante de força k e seu deslocamento x . (B) Análise de modo normal: aproximação harmônica da superfície de energia potencial. Para qualquer sistema biológico, a superfície de energia real é acidentada (linha pontilhada), mas para a análise de modo normal, a superfície é aproximada como uma superfície harmônica (linha simples). (C) Transformação de coordenadas cartesianas em coordenadas de modo normal. Neste exemplo de dois sistemas de coordenadas, as coordenadas do modo normal q são independentes, mas não as coordenadas cartesianas X . (D) Vetor de modo normal simples para a molécula de água. Cada seta representa a direção do movimento que cada átomo sofrerá conforme obtido da teoria do modo normal. O modo normal prevê três movimentos distintos para a molécula de água: modo simétrico, assimétrico e modo de flexão.

Fonte: adaptado de [57]

1.9 Aprendizado de Máquina

Recentemente a comunidade científica voltou sua atenção para uma área de conhecimento conhecida como Inteligência Artificial - IA. A IA veio ganhando força nos últimos tempos e pode ser aplicada de diversas formas. A IA foi estabelecida em 1950 pelo matemático Alan

Turing; ele estabeleceu como criar máquinas inteligentes e como testá-las. O teste de Turing ainda é usado atualmente. O termo Inteligência artificial foi cunhado oficialmente em 1956. A IA passou por um período de “escuridão” e vem ganhando força no cenário atual. Cenário este de densos conjuntos de dados, onde temos um excesso de informação e poder de Hardware para aplicar IA para as mais diversas áreas [65].

A IA pode ser dividida em alguns ramos e sub áreas, como o aprendizado de máquina. Os algoritmos de aprendizado de máquina (ML, do inglês *Machine learning*) possuem uma característica importante que é a capacidade distinta de aprender o ambiente circundante a partir de dados de entrada, ou seja, são capazes de treinar computadores para executar tarefas de forma inteligente além da análise de números tradicionais, aprendendo o ambiente circundante por meio de repetidos exemplos. Esta tecnologia foi aplicada em campos diversos como reconhecimento de padrões, visão computacional, engenharia de naves espaciais, entretenimento e aplicações biomédicas e médicas [67,68,69,71,73].

Dentro do aprendizado de máquina o método de seleção de atributos vem ganhando visibilidade em meio ao avanço da disponibilidade de dados. A seleção de atributos consiste em selecionar algumas características de um dado sistema, sendo estas selecionadas com critérios que as descrevem como melhores dentro daquele conjunto de dados. Sua função é redução da dimensionalidade dos dados, assim como melhorar a eficiência do aprendizado removendo dados redundantes e irrelevantes.

A seleção de atributos tem sido usada em muitos campos, como reconhecimento de imagem, recuperação de imagem, mineração e análise de dados bioinformáticos [74,75,75,77,78,79]. De acordo com os dados utilizados pode-se categorizar o método supervisionado, não supervisionado e parcialmente supervisionado. O método supervisionado leva em consideração a entrada de dados rotulados com etiquetas de identificação de acordo com o objetivo do aprendizado. No não supervisionado, os dados não são rotulados e o parcialmente supervisionado, os dados são parcialmente rotulados [80].

Alguns ambientes de código aberto foram disponibilizados pela comunidade de aprendizado de máquina. Esses ambientes possuem inúmeros algoritmos de aprendizados de máquina. O WEKA (do inglês *Waikato Environment for Knowledge Analysis*) é um desses ambientes, desenvolvido pela Universidade de Waikato. Nele podem se realizar vários estudos de aprendizado incluindo métodos como a seleção de atributos [37,66]. Nesse trabalho todas as etapas de aprendizado de máquina foram construídas com o auxílio do WEKA.

2. Objetivos

2.1 Objetivo geral

Identificar padrões estruturais, dinâmicos e/ou energéticos de resistência a herbicidas inibidores de Acetolactato Sintase (ALS) através de análise de comportamento dinâmico de ALSs resistentes.

2.2 Objetivos específicos

- Detectar resíduos descritores de mobilidade dos sistemas em dados de dinâmica molecular.
- Identificar potenciais redes de alosterismo envolvendo posições de mutação e possíveis resíduos descritores de movimento.
- Analisar a afinidade pelos herbicidas em sistemas resistentes e susceptíveis em relação à selvagem.
- Inferir sobre a separatividade dos sistemas resistentes e susceptíveis a herbicidas em termos de componentes principais amostrados durante a MD.
- Realizar análise farmacofórica dos contatos existentes em diferentes sistemas.
- Combinar os dados obtidos nos objetivos anteriores para análise de seleção de atributos por aprendizado de máquina, verificando padrões de resistência em IMI e SU.

3. Materiais e métodos

3.1 Coleta de dados/Modelagem dos sistemas

Estruturas tridimensionais de Acetolactato sintase se encontram disponíveis na base de dados PDB [34] (www.rcsb.org/). Dentre as estruturas, até o momento disponíveis, quatorze são oriundas do organismo modelo *A. thaliana*. Para a realização deste trabalho usaram-se alguns critérios de seleção para a escolha das melhores estruturas de acordo com a intenção do trabalho.

O primeiro critério foi a presença de herbicidas da classe IMI ou SU, ou a ausência de qualquer molécula inibidora (APO). Apenas uma estrutura se encontra disponível com a presença do inibidor IMI e uma sem qualquer inibidor. Dentre as estruturas com presença de SU, um segundo critério foi aplicado, sendo a presença do cofator ThDP em sua forma intermediária e com boa resolução (considerando o b-fator), posto ter sido inicialmente notada uma variância quanto a estes atributos dentre as estruturas disponíveis com esta classe de ligantes. Foram também avaliados os valores de B-factor, ausência de átomos, dupla ocupância, e resolução do cristal.

Após a aplicação dos critérios o pdb APO selecionado foi o 5K,Q o pdb com presença de IMI, o 1z8n e o pdb com presença de SU, o 3EA4. O cofator ThDP encontra-se em seu estado intermediário HETHDP (TDM) nos PBDs 5K6Q e 3EA4 e não no PBD 1Z8N. Para igualarmos os cofatores, e assim obter resultados com menos artefatos, acrescentou-se através do *software* AVOGADRO o acetato faltante no ThDP do modelo gerado para o pdb 1Z8N.

Um tratamento prévio às simulações foi realizado. Retirou-se manualmente átomos com dupla ocupação, considerando os átomos com maior ocupação. Para um trabalho mais realista obtive o dímero cristalográfico presente nos pbs (o sítio ativo se encontra na interface do dímero). Artefatos cristalográficos como NHE (tampão) foram retirados e os cofatores e herbicidas foram mantidos. As águas em proximidade de até 7 Å dos cofatores e proteínas foram mantidas. Em seguida, para inserção dos átomos faltantes realizou-se um procedimento de modelagem através do *software* Swiss-model inserindo como template o dímero cristalográfico gerado no pymol. O modelo e template obtiveram RMSDs respectivamente iguais a 0,055 Å (5k6q), 0,080 Å (1z8n) e 0,050 Å (3ea4).

Para construção dos mutantes o software Modeller foi utilizado. Os modelos gerados pelo Swiss model para cada sistema com inibidor (IMI-1z8n e SU-3ea4) foram mutados de acordo com a **Tabela 1**. Para cada modelo com inibidor foi construído um modelo com cada uma das mutações, totalizando 22 mutantes.

Tabela 1. Mutações inseridas nos modelos com herbicidas. Numeração dos resíduos de acordo com ALS de *Arabidopsisthaliana*.

	Mutações	Resistência a IMI	Resistência a SU	Referência
1	A122T	X		86
2	P197A		X	87
3	P197W		X	88
4	P197W_D376E	X	X	88,89
5	A205V	X		90
6	E268G		X	88
7	D376E	X	X	89
8	R377H		X	91
9	K434N		X	88
10	W574L	X	X	92
11	S653N	X		93

3.2 Simulações de MD e Simulações de metadinâmica com escape do ligante

O estado de protonação dos resíduos foi calculado através do *software* H++, assumindo condições fisiológicas de salinidade (0,15 M), pH 7,0 e constantes dielétricas relativas internas e externas de 10 e 80 respectivamente. As cargas dos átomos dos ligantes e dos cofatores foram calculadas usando o método AM1-BCC (FAD carga -2, TDM carga -3, 2SM carga 0, 1IQ carga -1) e os parâmetros destes foram gerados usando o campo de força geral do Amber (GAFF, do inglês General Amber Force Field) utilizando o módulo ANTECHAMBER do AMBERTools 18 [40].

As estruturas foram processadas com o programa Tleap do AmberTools 18, atribuindo então os parâmetros de mecânica molecular. Os átomos das proteínas, moléculas de água e íons foram representados com a versão ff14SB do campo de força AMBER, com uma caixa de água octaédrica que se estendeu em 12 Å além das maiores dimensões proteicas em cada eixo X, Y e Z, e a representação TIP3P foi usada para as moléculas de água. Íons Na⁺ e Cl⁻ foram adicionados para neutralização e força iônica de 0,15 M.

As simulações foram realizadas utilizando o *software* NAMD 2.13, com um passo de integração de 2fs. Temperatura e pressão foram controladas pelo algoritmo de Langevin em

300K e 1 atm respectivamente. Condições periódicas de contorno foram utilizadas. Para interações eletrostáticas de longo alcance utilizou-se Particle Mesh Ewald (PME).

O protocolo de relaxamento consistiu em:

- 20000 passos de minimização por gradiente conjugado;
- 500 ps de MD com restrições harmônicas para todos os átomos da proteína, cofatores e ligante.
- 200 ps de MD com restrições harmônicas para todos os átomos da proteína, ligantes e cofatores, exceto C27, C29, O28 do ThDP (átomos acetado inseridos para que todos os sistemas estivessem com o ThDP em seu estado intermediário).
- 200 ps de MD com restrições harmônicas para todos os átomos da proteína.
- 200 ps de MD com restrições harmônicas para todos os átomos da proteína, exceto para cadeias laterais especiais (átomos faltantes acrescentados no Swiss model e átomos dos resíduos inseridos nas mutações)
- 500 ps de MD com restrições harmônicas apenas para os átomos do backbone.
- 500 ps de MD sem restrições harmônicas.
- 12 ns de MD de produção sem restrições harmônicas e com reinicialização anterior das velocidades de acordo com temperatura e pressão 300K e 1 atm, respectivamente.

Apenas para quatro sistemas onde a mutação era distante dos sítios a última etapa de 12 ns foi estendida para 40ns. Isto foi feito de forma a melhor amostrar os potenciais desvios na estatística conformacional relevante ao ambiente do ligante (sítio ativo) decorrentes de tais substituições significativamente distantes desta mesma região. Os sistemas em tal situação foram os mutantes E268G e K434N, em ambos os casos, tanto para o sistema com associação do ligante IMI quanto com associação de SU. Foram removidos os primeiros 10 ns das trajetórias finais desses sistemas. Os 30 ns restantes foram divididos em duas trajetórias, obtendo assim trajetórias com aproximadamente o mesmo número de frames. Seguiu-se as análises posteriores com mais esses sistemas, totalizando 29 sistemas (2 wt/holo, 1 wt/apo e 26 holo/mutantes).

Para simulações de metadinâmica, foram utilizadas como variáveis coletivas a distância de átomos centrais de cada ligante (o átomo C9 do ligante 2SM da classe SU e o átomo C2 do ligante 1IQ da classe das IMI) em relação ao cofator ThDP (variável CV1, com 0 a 40 Å de abrangência), bem como o ângulo compreendendo estes dois centros e o centro geométrico do ligante como um todo (variável CV2 com abrangência entre -180 a 180°. As *grids* da variável CV1 tiveram dimensões de 1Å enquanto as da variável CV2 de 10 °. A altura do potencial Gaussiano foi de 0,05 kcal/mol e sua largura de 1,77 grids para cada respectiva variável

coletiva, sendo tal potencial adicionado a cada 2ps. A conformação oriunda da última etapa do procedimento de equilíbrio nas simulação de MD foram utilizadas como ponto inicial para a simulação de metadinâmica foi escolhida através de uma análise de PCA. Essa análise foi aplicada sobre os resíduos em distância de contato (7 Å) em relação ao ligante e pertencentes aos loops na abertura do sítio ativo. A partir das projeções nas componentes que mais abundantemente descreviam a abertura do sítio e de espaço para a saída do ligante, foi escolhido a configuração cuja projeção demonstrava maior abertura do sítio, permitindo um caminho de saída para o ligante. Isto foi feito para evitar eventuais artefatos nas superfícies de energia livre obtidas nos procedimentos de metadinâmica decorrentes de acúmulo de barreiras cinéticas ao longo do caminho de saída dos ligantes.

Os protocolos de metadinâmica foram realizados com conformação fixa do receptor por restrição harmônica, excetuando os resíduos determinantes da abertura do caminho de saída do ligantes previamente determinados no protocolo de PCA citado no parágrafo anterior. A reintrodução das restrições harmônicas nas regiões pertinentes do receptor preparatoriamente à metadinâmica foi efetuada a partir do seguinte protocolo gradual:

- 500 ps de MD reintroduzindo as restrições harmônicas para todos os átomos da proteína, cofatores e do ligante.
- 200 ps de MD com remoção das restrições harmônicas das cadeias laterais dos resíduos bloqueadores da saída do ligante.
- 200 ps de MD com remoção das restrições harmônicas do backbone e cadeias laterais dos resíduos bloqueadores da saída do ligante.
- 300 ps de MD com remoção das restrições harmônicas também do ligante.
- 1 ns de MD nas condições finais acima mencionadas com reinicialização das velocidades de acordo com temperatura e pressão 300K e 1 atm, respectivamente.

O protocolo de metadinâmica propriamente dito foi efetuado com o sistema previamente preparado conforme o acima exposto, constando de aproximadamente 5 ns (tempo aproximado no qual se verificou, para todos os sistemas a saída e o *re-crossing* do ligante), em protocolo similar ao efetuado em [94].

3.3 Predição de resíduos descritores por aprendizado de máquina

Usamos técnicas de seleção de atributos atreladas aos valores de RMSF de cada resíduo recuperados das simulações de dinâmica molecular para predizer resíduos descritores de movimento nas ALSs. As trajetórias foram divididas em dois grupos: proteínas ligadas a IMI e proteínas ligadas a SU. Dentro de cada grupo as trajetórias de cada sistema foram subdivididas

em duas classes, proteínas resistentes ou susceptíveis ao ligante em questão. Usou-se o pacote para aprendizado de máquina WEKA, com a função de seleção de atributos *CfsSubsetEval* e método *BestFirst* [37,66], para selecionar os resíduos que colaboram para a separação dos sistemas em uma certa classe (susceptíveis e resistentes). Essa função avalia o valor de um subconjunto de atributos, considerando a capacidade preditiva individual de cada recurso junto com o grau de redundância entre eles. Os valores de RMSF para cada resíduo foram utilizados como input para a seleção de atributos, sendo cada sistema uma instância.

Após a seleção dos resíduos descritores de IMI e de SU, um classificador baseado em múltiplas camadas de percéptrons (MLP - *multilayer perceptron*) foi construído (através do WEKA) para IMI e outro para SU com seus respectivos resíduos descritores e valores de RMSF. O critério de validação cruzada foi acionado com 3 *folds*. Os sistemas foram considerados as instâncias (identificados como sistemas resistentes ou susceptíveis), e os atributos os valores de RMSF.

3.4 Redes Alostéricas

Prody [38] é uma API integrada desenvolvida para modelar e analisar a dinâmica de proteínas, é um pacote Python gratuito e de código aberto. Cálculos de PRS- *Perturbation Response Scanning* podem ser feitos através do Prody por modelos de redes anisotrópicas (ANM). Cada resíduo possui outros resíduos conectados a ele, sendo estes os sensores (recebem os efeitos estruturais causados pela posição) ou efetores (provocam mudanças estruturais na posição em questão). Submeteu-se cada resíduo de mutação da estrutura WT/APO ao PRS e a partir dos resíduos retornados, identificados como efetores ou sensores, construiu-se uma rede de alosterismo para cada posição de mutação. Essa rede foi construída através de um script python e software de visualização molecular Pymol [39].

3.5 Análise de modos normais de Componente principal (PCA)

Utilizou-se o ProDy para construção dos modos normais (modelo de rede anisotrópica- ANM) através da proteína modelada em dímero sem ligantes (WT/APO). Os PCs mais representativos da dinâmica desse sistema também foram calculados pelo ProDy. Três análises iniciais de PCA foram realizadas levando em consideração: i) todos os CA, ii) CA de resíduos descritores de IMI e iii) CA de descritores de SU. Matrizes de sobreposição entre os modos normais e PCs foram construídas pela função *calcOverlap()* do ProDy. Os autovetores e autovalores

associados aos CA descritores na análise de modos normais foram retirados manualmente dos arquivos de saída.

Outra análise de PCA foi realizada para cada sistema com IMI ou SU, levando em consideração os CA de descritores IMI e SU, respectivamente. Novas matrizes de sobreposição foram construídas.

Uma análise de PCA sobre dois conjuntos de trajetórias: i) trajetória WT/APO, WT/holo, holo/mutantes com IMI (levando em consideração CA dos descritores), ii) trajetória WT/APO, WT/holo, holo/mutantes com SU (levando em consideração CA dos descritores), foi realizada. Novas matrizes de sobreposição foram construídas.

Para análises de PCA, trajetórias de dinâmica molecular dos átomos correspondentes foram extraídas usando o programa CPPTRAJ e as análises foram realizadas para os últimos 8 ns de dinâmica de produção (5–12 ns) usando 2000 frames de cada sistema. Para as análises de conjuntos de trajetória, o total de 58.000 frames foi obtido, sendo usados 7.250 frames (250 frames de cada sistema).

3.6 Análise de afinidade pelo ligante

Após a realização das simulações de metadinâmica, para selecionar os frames/estruturas a serem considerados ao longo da reconstrução PMF, utilizou-se como métrica o momento em que os dois mínimos de energia pertinentes (mínimo do ligante no interior do sítio e mínimo do ligante na água) terem sido amostrados. Sendo selecionado o frame imediatamente anterior ao evento considerado *recrossing* ou retorno a um desses dois mínimos, em protocolo similar ao efetuado em [94]. Para isso, avaliou-se: i) se a energia de VdW nessa etapa estava próxima de 0 (sem interação com a proteína), se a distância do ligante em relação ao ThDP ainda não havia sido amostrada nas bordas na proteína e se o ligante estava imerso na água e não fazendo contato com a proteína. Assim que o ligante explora o ambiente da água e antes de iniciar o seu recrossing (voltar a fazer contato com a proteína) o arquivo pmf salvo até essa etapa da simulação é selecionado para reconstrução dos perfis de energia livre (free energy landscape - FEL) e cálculos de afinidade do ligante pela proteína. Esses critérios evitam o sobrepreenchimento dos mínimos de energia com os potenciais gaussianos.

Um script interno em python é rodado afim de se encontrar os mínimos de energia e então calcular os valores de ΔG entre o estado ligado à proteína e o estado livre (na água). Para comparação dos sistemas mutantes (holo/mutantes) e sistemas sem mutação (WT/holo) calcula-se o $\Delta\Delta G$, ΔG do sistema mutante diminuindo o ΔG do seu respectivo WT ($\Delta\Delta G = \Delta G$ holo/mutante - ΔG WT/holo).

Uma segunda análise de afinidade foi construída através do programa MM/PBSA.py [41] disponível no AmberTools. O método Poisson Boltzmann(PB) foi utilizada para a porção eletrostática. Os valores de contribuição de energia para o complexo (proteína com ligação nos dois sítios), para o receptor (proteína sem os ligantes) e para cada um dos ligantes foram recolhidos.

Para análise de entropia quase harmônica, o programa Wordom [42] foi utilizado. Para cada valor de contribuição de energia do MM/PBSA foi adicionada a aproximação quase harmônica. Através destes dados calculou-se a afinidade por diferentes abordagens: i) ΔG do complexo ternário, ii) ΔG de associação do segundo ligante e ii) $\Delta(\Delta G)$ de cooperatividade entre os ligantes. Lembrando que para efeito de comparação, o cálculo de $\Delta\Delta G = \Delta G$ holo/mutante - ΔG WT/holo foi realizado.

ΔG do complexo ternário pode ser representado pela **Equação 3**:

$$\Delta G_t = \left(G_{\text{complexo12}} - (G_{\text{receptor}} + G_{\text{ligante1}} + G_{\text{ligante2}}) \right) - (Sh0_{\text{ligante1}} - Sh0_{\text{ligante2}} + Eh0_{\text{liagnte1}} + Eh0_{\text{ligante2}})$$

Equação 3

onde, $G_{\text{complexo12}}$ indica energia do complexo quando os dois ligantes estão ligados, G_{receptor} energia do receptor (proteína sem ligantes), G_{ligante1} energia do primeiro ligante e G_{ligante2} energia do segundo ligante. A parte quase harmônica é indicada por: $Sh0_{\text{ligante1}}$ entropia vibracional do primeiro ligante, $Sh0_{\text{ligante2}}$ entropia vibracional do segundo ligante, $Eh0_{\text{ligante1}}$ energia vibracional do primeiro ligante e $Eh0_{\text{ligante2}}$ energia vibracional do segundo ligante.

ΔG de associação do segundo ligante foi calculado com base na **Equação 4**:

$$\Delta G_{2\text{ligante}} = -RT \ln \left(\frac{\exp\left(-\frac{G_{\text{complexo12}}}{RT}\right)}{\exp\left(-\frac{G_{\text{complexo1}} + G_{\text{ligante2}}}{RT}\right) + \exp\left(-\frac{G_{\text{complexo2}} + G_{\text{ligante1}}}{RT}\right)} \right)$$

Equação 4

onde, R é constante de Boltzmann, T a temperatura, $G_{\text{complexo12}}$ é a energia do complexo quando os dois ligantes estão ligados (neste caso já somando a parte vibracional dos ligantes), o $G_{\text{Complexo1}}$ é a energia do complexo com apenas um ligante (já com sua parte entrópica), G_{ligante2} energia apenas do ligante 2, $G_{\text{Complexo2}}$ energia do complexo com apenas o ligante 2 ligado (já com sua parte entrópica) e G_{ligante1} energia apenas do ligante 1.

A equação que representa o $\Delta(\Delta G)$ de cooperatividade é a **Equação 5**:

$$\Delta(\Delta G) = \Delta G_t + 2(RT \ln \left[\exp\left(-\frac{\Delta G_{\text{ligante1}}}{RT}\right) + \exp\left(-\frac{\Delta G_{\text{ligante2}}}{RT}\right) \right])$$

Equação 5

onde, ΔG_t é o ΔG do complexo ternário calculado na equação 1, $\Delta G_{\text{ligante1}}$ é o valor de Δ da energia do ligante 1 (já com sua parte entrópica) e $\Delta G_{\text{ligante2}}$ é o valor de Δ da energia do ligante 2 (já com sua parte entrópica).

3.7 Reconstrução dos perfis de energia livre e PCA

Foi realizada a reconstrução dos perfis de energia livres para cada sistema (WT/APO, WT/holos e holos/mutantes) através do kit de ferramentas PyRewighting2D [43,44]. Considerou-se as projeções nos três primeiros autovetores, PC1, PC2, PC3. A projeção média de cada sistema foi projetada para uma melhor visualização de deslocamento das diferenças estruturais em cada sistema. A estrutura inicial usada na metadinâmica também foi projetada nos perfis.

Uma segunda reconstrução dos perfis de energia livre foi realizada sobre dois conjuntos de trajetórias: trajetórias de sistemas com IMI e trajetórias de sistemas com SU. Projetou-se sobre essa reconstrução as estruturas com projeções médias de cada sistema e também as estruturas iniciais da metadinâmica de cada sistema em busca de um padrão de mobilidade dentro do espaço PC1x PC2 e também PC3.

3.8 Análise Farmacofórica

Em uma busca por padrões de contato atômico nas estruturas projetadas sobre a reconstrução dos perfis de energia livre dos conjuntos de trajetória mencionado no item anterior, o webservice nAPOLI [36] foi utilizado. Dentre os sistemas previamente classificados como resistente e susceptível, foi utilizado o webservice nAPOLI para analisar os contatos entre a proteína e o ligante e a proteína e cofatores. Tais contatos foram levados ao WEKA para de seleção de atributos (contatos), de forma a busca padrões de contatos que separassem as classes resistentes das susceptíveis. Os contatos atômicos entre os cofatores, herbicidas e proteína foram utilizados como entrada para a seleção de atributos, sendo cada sistema uma instância.

Os contatos retornados na seleção de atributos anterior foram levados para um classificador MLP construído ainda no WEKA para IMI e outro para SU. O critério de validação cruzada foi acionado com 3 *folds*.

3.9 Aprendizado de máquina e dados de MD

Todos os experimentos computacionais realizados com o aprendizado de máquina foram implementados no WEKA, com a função de seleção de atributos *CfsSubsetEval* e método *BestFirst*. Os sistemas foram rotulados previamente em duas classes: resistentes e susceptíveis. Três análises de seleção de atributos foram realizadas:

- i. seleção de resíduos descritores já detalhada no item 3.3.
- ii. seleção de contatos importantes retornados no item 3.8.
- iii. seleção de descritores gerais: PCs (retornados no item 3.7) e $\Delta\Delta$ Gs (retornados no item 3.6).

Após a seleção de atributos, para avaliação dessa seleção um classificador *multilayer perceptron* (MLP) sempre foi construído. Em todas as análises houve aplicação da técnica para o conjunto de sistemas com IMI e outra aplicação para sistemas com SU. O critério de validação cruzada foi acionado com 3 folds em todas as análises.

4. Resultados e discussões

4.1 Análises de RMSD e RMSF das trajetórias de MD

Após a realização de simulações de dinâmica molecular de 12ns de cada sistema, a análise de RMSD promovida contra o primeiro frame de cada trajetória, considerando apenas o backbone da proteína, sugere uma tendência ao equilíbrio. Assim obtendo uma boa amostragem para as análises posteriores (**Figura 8**).

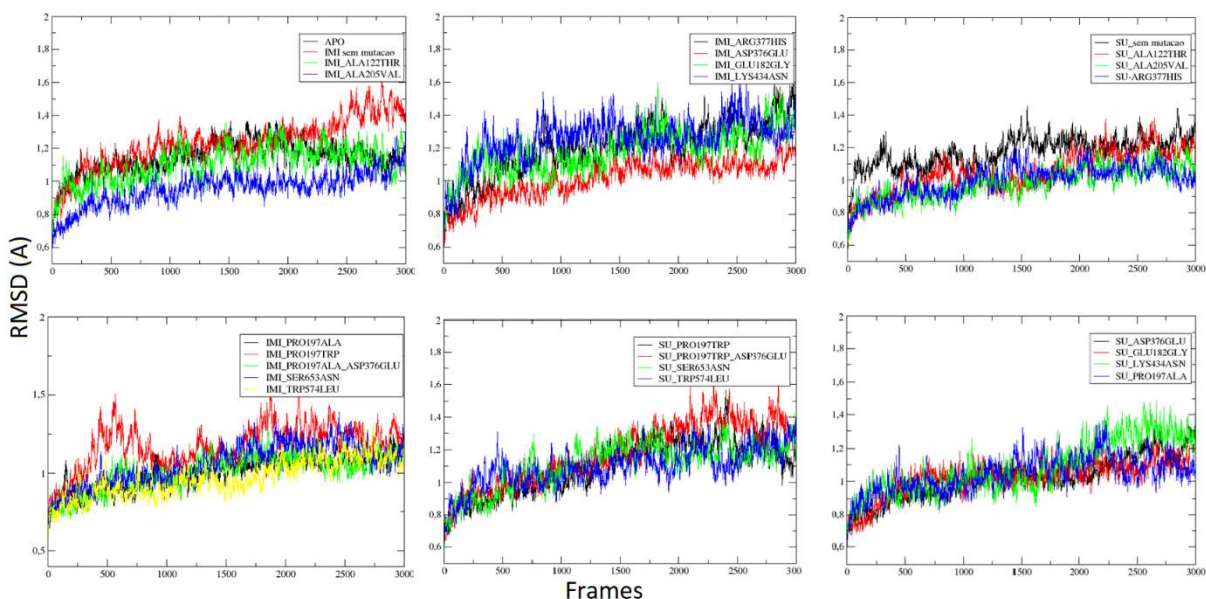


Figura 8.RMSD ao longo da trajetória de cada simulação de MD relacionada ao seu respectivo frame inicial. Fonte: próprio autor.

A análise de RMSF (**Figura 9**) de cada trajetória foi realizada a fim de se identificar resíduos importantes para a mobilidade da proteína. As análises visuais foram descartadas pelo fato de se ter um grande volume de gráficos para analisar e então a enorme dificuldade em identificar esses resíduos, considerando todos os sistemas. Portanto, essa análise foi direcionada a uma análise mais sistemática e preditiva de resíduos descritores por aprendizado de máquina.

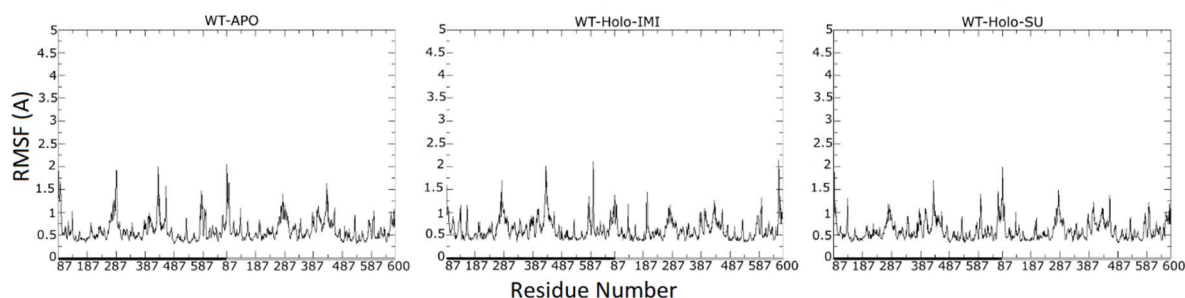


Figura 9.RMSF ao longo da trajetória de MD, gráficos de sistemas mutantes não apresentados na figura.Cadeia A e B separadas no eixo x por barra preta e cinza, respectivamente. Fonte: próprio autor.

4.2 Aprendizado de máquina e resíduos descritores de mobilidade

A função de seleção de atributos *CfsSubsetEval*, implementada no WEKA nos retornou em uma primeira análise 139 resíduos descritores de movimento com base em dados de RMSF das trajetórias de MD de enzimas ligadas ao herbicida IMI (WT/holo e holo/mutantes). Esses resíduos foram avaliados manualmente, verificando quais contribuem mais para a separação dos sistemas em resistentes e susceptíveis. Após a aplicação dessa abordagem manual, apenas 24 resíduos foram selecionados. No caso das trajetórias com ligante SU (WT/holo e holo/mutantes), a abordagem manual não foi necessária, a seleção de atributos nos retornou apenas 35 resíduos importantes para a mobilidade.

Para verificação da contribuição desses resíduos descritores para classificação em resistentes ou susceptíveis, um classificador MLP foi construído no WEKA, para os sistemas com IMI e outro para os sistemas com SU. Os resultados dos classificadores foram excelentes, com uma acurácia de 1 e área sob a curva ROC de 1, demonstrando que os resíduos selecionados conseguem descrever movimentos que diferenciam dinâmicas de proteínas inibidas (susceptíveis) de dinâmicas de proteínas não inibidas (resistentes) (**Tabela 2**).

Direcionando as análises para esses resíduos, e verificando estruturalmente a localização dos mesmos, percebe-se uma diferença entre os descritores dos sistemas com IMI e SU. Estes resíduos são diferentes, distribuídos de maneira diferente sobre a superfície da proteína (**Figura 10 e Figura 11**). Os 24 descritores de sistemas com IMI se encontram distribuídos por toda a extensão da proteína, mais localizados em um monômero. No caso dos descritores de SU, esses se encontram próximos aos sítios ativos, em ambos os monômeros.

Tabela 2. Resultados obtidos para o classificador MLP de IMI e de SU.

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	res
	1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	sus
Weighted Avg.	1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	

Alguns estudos sobre enzimas ativas em sua forma homodimerizada, trazem insights sobre o comportamento dinâmico diferenciado entre os monômeros. Suarez e Dias [35] realizaram simulações dinâmicas para avaliar o comportamento da protease 3CL^{pro} do corona vírus denominado SARS-CoV-2 (Severe acute respiratory syndrome coronavirus 2). Neste

trabalho eles demonstraram que a movimentação dos resíduos da alça de oxianion foi diferenciada para os monômeros, sendo que um se encontra mais rígido que o outro sem comprometer a acessibilidade dos locais de ligação importantes. O mesmo parece ocorrer em sistemas ligados ao herbicida imidazolinona.

Esses dados nos permitem hipotetizar que os resíduos envolvidos nas mudanças estruturais advindas das mutações pontuais (que permitiram a proteína estar ou não inibida pelos ligantes), são diferentes para herbicidas diferentes, mesmo em mutações que causam resistência a ambos os herbicidas, sendo que com SU os resíduos descritores de movimento permeiam mais o sítio ativo que os descritores de IMI.

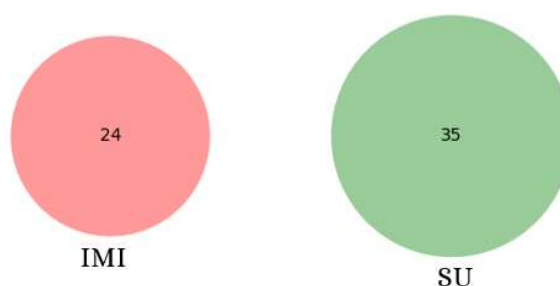


Figura 10. Diagrama de Venn entre os resíduos descritores de IMI e de SU. Não existem o compartilhamento dos resíduos descritores entre IMI e SU.

Fonte: próprio autor.

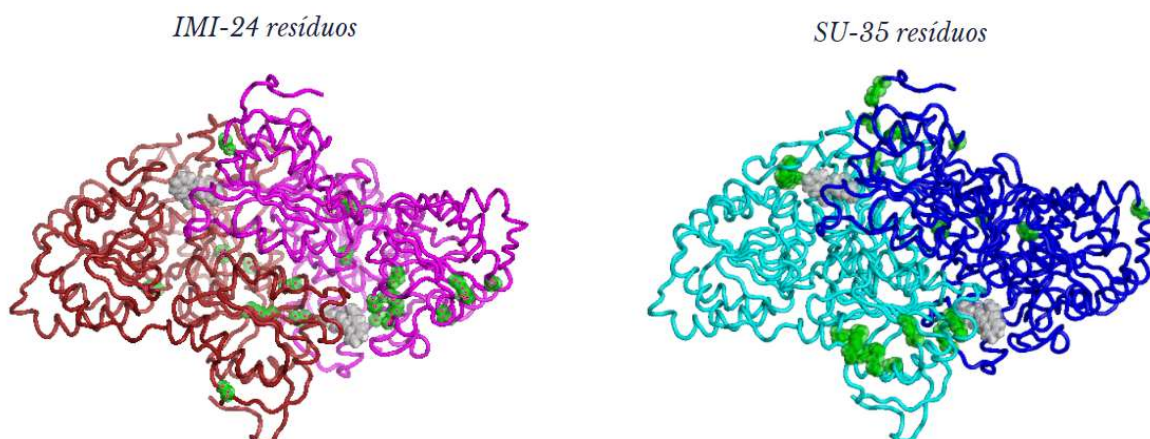


Figura 11. Localização dos resíduos descritores de IMI e SU na estrutura tridimensional da proteína. Monômeros coloridos de cores diferentes. Resíduos descritores encontrados na seleção de atributos em esferas verdes. Herbicidas em esferas cinza.

Fonte: Laura Magalhães.

4.3 Redes alostéricas

Resíduos *efetores* são aqueles cuja perturbação tem grandes efeitos na estrutura e na dinâmica, enquanto que resíduos *sensores* respondem mais fortemente às perturbações dos efetores e eles

próprios sofrem mudanças estruturais. A análise de *Perturbation Response Scanning* (PRS) sobre cada posição de mutação foi capaz de nos retornar resíduos sensores e efetores das mesmas. A **Figura 12** mostra a rede de alosterismo (efetores e sensores) de cada posição de mutação e também a localização dos resíduos descritores de IMI e SU.

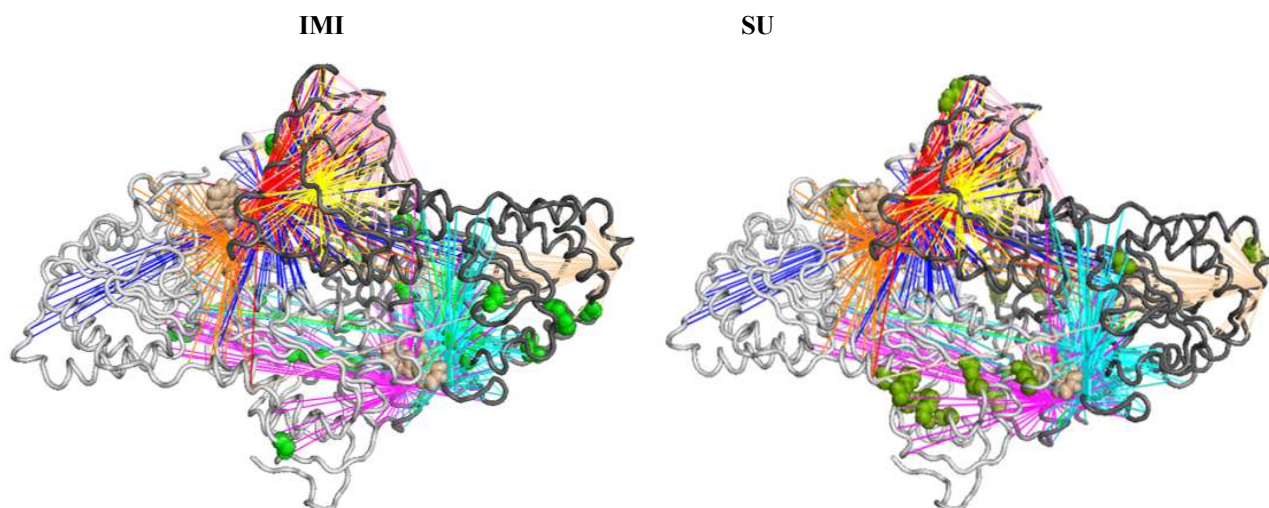


Figura 12. Rede alostérica e localização dos resíduos descritores de IMI e SU na estrutura tridimensional da proteína. Resíduos descritores em esferas verdes. Arestas coloridas (uma cor para cada posição de mutação) representam ligação entre a posição de mutação (saída das arestas) e os resíduos sensores/efetores dessa posição (direção das arestas).

Fonte: próprio autor.

Cada conjunto de cores de arestas representa uma posição de mutação relatada na literatura como resistente ou susceptível a IMI e SU. As arestas partem da posição de mutação em direção a seus resíduos efetores ou sensores. Analisando a rede construída pode-se notar que as arestas parecem partir das regiões onde se encontram os resíduos descritores (em esferas verdes), ou se encaminham para regiões onde esses resíduos estão localizados, sugerindo que as posições envolvendo mutações que alteram a sensibilidade ao ligante estão envolvidas em regiões alostéricas.

4.4 Análise de sobreposição de PCA e modos normais

Inicialmente os modos normais usando o modelo de rede anisotrópica - ANM foram calculados para a proteína dimerizada e estática sem ligante (WT/APO), assim como os cálculos de PCA que foram feitos sobre a dinâmica molecular do sistema WT/APO. Tanto para o ANM quanto para o PCA levou-se em consideração: i) todos os carbonos alfa de todos os resíduos, ii) carbonos alfa dos resíduos descritores de IMI e iii) carbonos alfa dos resíduos descritores de SU. Para verificação da similaridade entre os modos normais calculados para a proteína estática

e os PCs calculados a partir da MD, construiu-se uma matriz de sobreposição (**Figura 13**) entre os trinta primeiros PCs e trinta primeiros modos normais. Matrizes mais diagonais, com valores de sobreposição mais altos (observados por quadrantes coloridos por cor mais vermelha) possuem valores de sobreposição mais elevados entre movimentos amostrados pela MD e movimentos intrínsecos da proteína calculados pelos modos normais.

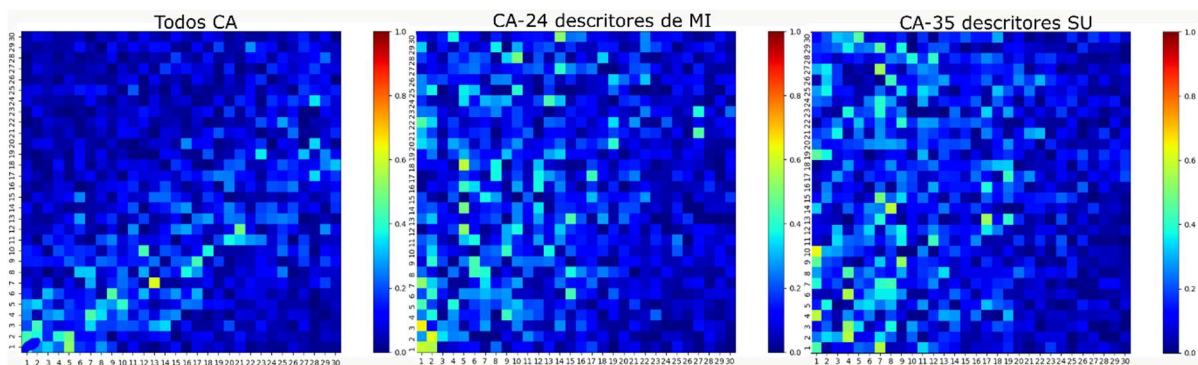


Figura 13. Matrizes de sobreposição entre os trinta primeiros PCs da trajetória WT/APO e trinta primeiros modos normais do primeiro frame da trajetória WT-APO.

Fonte: próprio autor.

Seguindo o preceito de que matrizes mais diagonais são aquelas em que o determinante da matriz é igual ao produto da sua diagonal, uma análise de diagonalidade foi realizada usando a fórmula descrita como:

$$Fuga\ diagonal = |Determinante - produto\ da\ diagonal|$$

Equação 6

A análise aponta para uma maior diagonalidade das matrizes (barras mais próximas de 0) quando estas foram produzidas com os resíduos descritores de IMI e SU em comparação com todos os resíduos da proteína (**Figura 14**). Resíduos descritores de dinâmicas de sistemas com IMI parecem estar mais próximos da dinâmica intrínseca da proteína.

Análises de autovalores da PCA para os 24 ou os 35 resíduos descritores e ANM de todos os CA permitem inferir uma regularidade nos dados, sendo estes confiáveis (**Figura 15** e **Figura 16**). Comparando-se os autovalores de PCA dos resíduos descritores, percebe-se que há uma maior magnitude dos movimentos dos descritores de IMI em relação aos de SU.

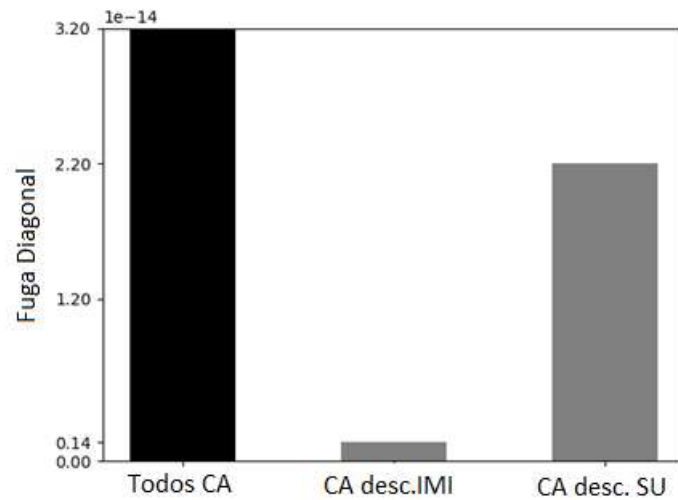


Figura 14. Gráfico de barras para valores de Fuga diagonal da análise de sobreposição entre os trinta primeiros PCs e trinta modos normais para sistema WT-APO.

Fonte: próprio autor.

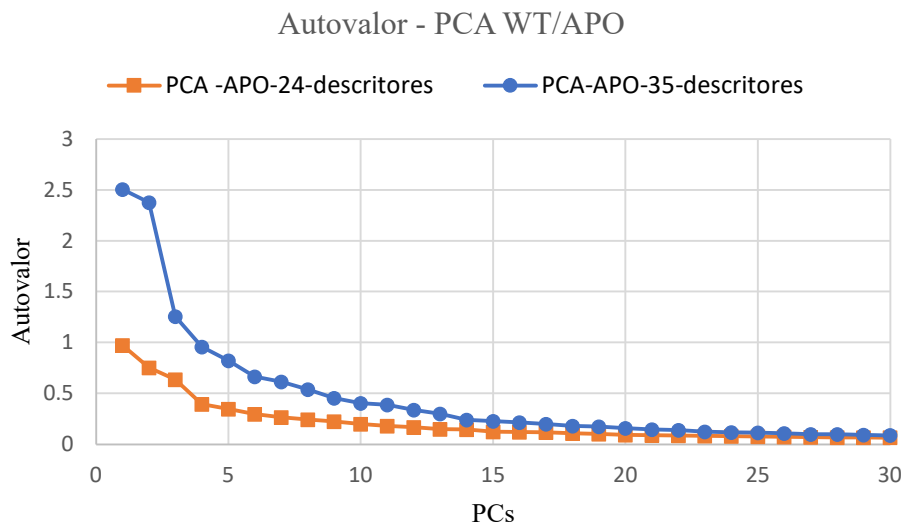


Figura 15. Autovalor das análises de PCA para o sistema WT/APO considerando os CA dos 24 descritores de IMI e dos 35 de SU.

Fonte: próprio autor.

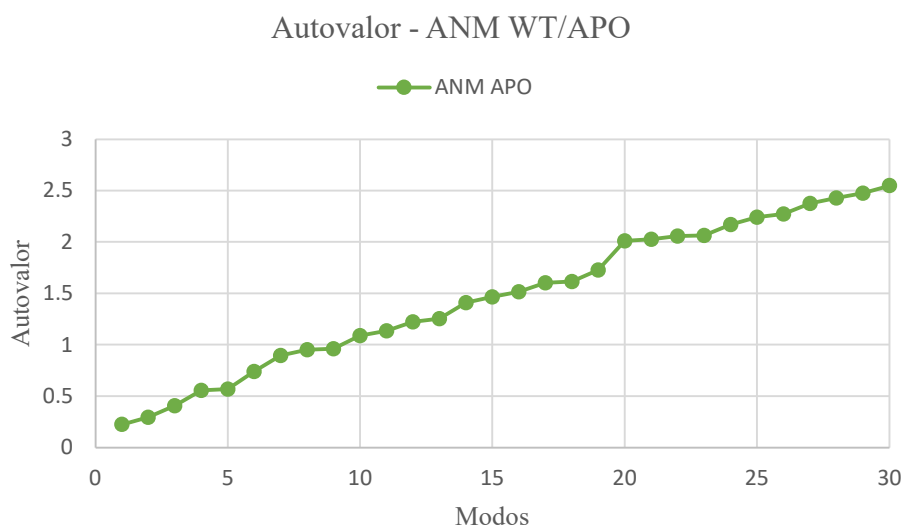


Figura 16. Autovalor da análise de modos normais para o sistema WT/APO considerando todos os CA. Fonte: próprio autor.

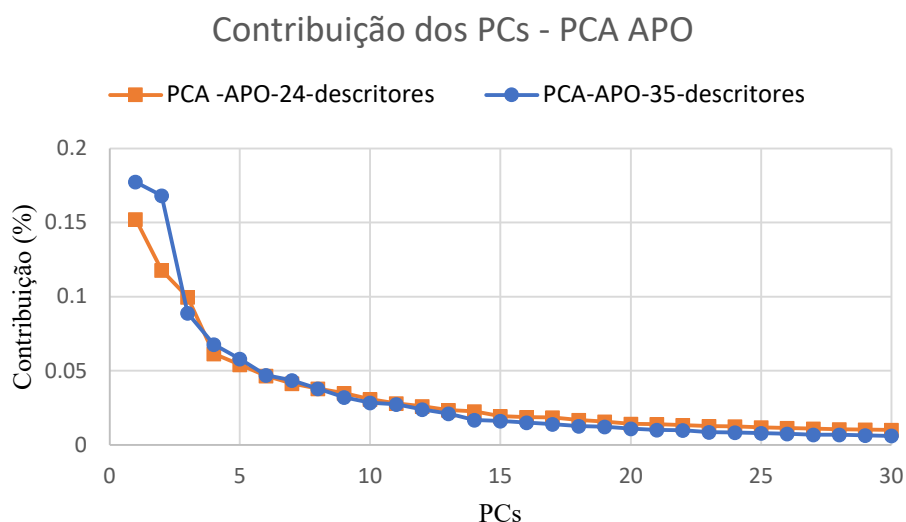


Figura 17. Contribuição de cada PC das análises de PCA para o sistema WT/APO considerando os CA dos 24 descritores de IMI e dos 35 de SU. Fonte: próprio autor.

Ainda verificando a similaridade entre os modos normais e os PCs, uma análise de sobreposição, agora de forma cumulativa, entre os três primeiros PCs e três primeiros modos normais foi realizada. A análise nos permite inferir o quão bem o espaço definido por um determinado conjunto de PCs pode representar um dado movimento descrito pelos modos normais.

A **Figura 18** mostra a sobreposição cumulativa entre os três primeiros PCs e três primeiros modos normais considerando-se os mesmos grupos de carbonos alfa descritos

anteriormente: i) todos os carbonos alfa de todos os resíduos, ii) carbonos alfa dos resíduos descritores de IMI e iii) carbonos alfa dos resíduos descritores de SU. Nota-se que existe uma maior sobreposição (valores mais elevados) relacionada aos resíduos descritores de IMI, corroborando a análise de sobreposição anterior.

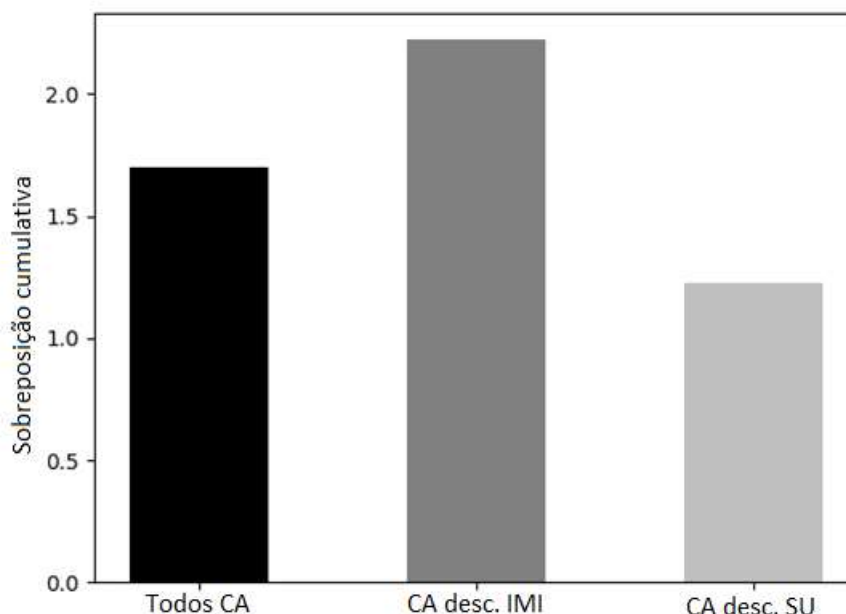


Figura 18. Gráfico de barras para valores de sobreposição cumulativa entre os três primeiros PCs e modos normais para sistema WT-APO.

Fonte: próprio autor.

Para uma melhor compreensão sobre os achados, matrizes de sobreposição entre os 30 primeiros PCs e 30 primeiros modos normais, foram construídas para todos os sistemas (WT/holos e holos/mutantes) considerando apenas os CA dos resíduos descritores e também para o conjunto de trajetória de todos sistemas IMI (WT/APO, WT/holo IMI e Holos/Mutantes com IMI) e para o conjunto de trajetória de todos os sistemas SU (WT/APO, WT/holo SU e Holos/Mutantes com SU). Quando se compara os descritores de IMI e de SU, as matrizes com descritores de IMI são mais próximas da diagonalidade (valores de Fuga diagonal mais próximos de 0), ou seja, os modos normais, movimentos intrínsecos da proteína são mais amostrados na MD de sistemas com IMI do que os com SU, mostrando também que estes movimentos são diferentes (**Figura 19**). Os movimentos das proteínas com IMI seguem os modos intrínsecos e se desviam deles apenas em intensidade de movimento, enquanto que os movimentos de proteínas com SU estão menos ligados aos movimentos intrínsecos e sim a movimentos diferentes dos preditos pelos modos normais.

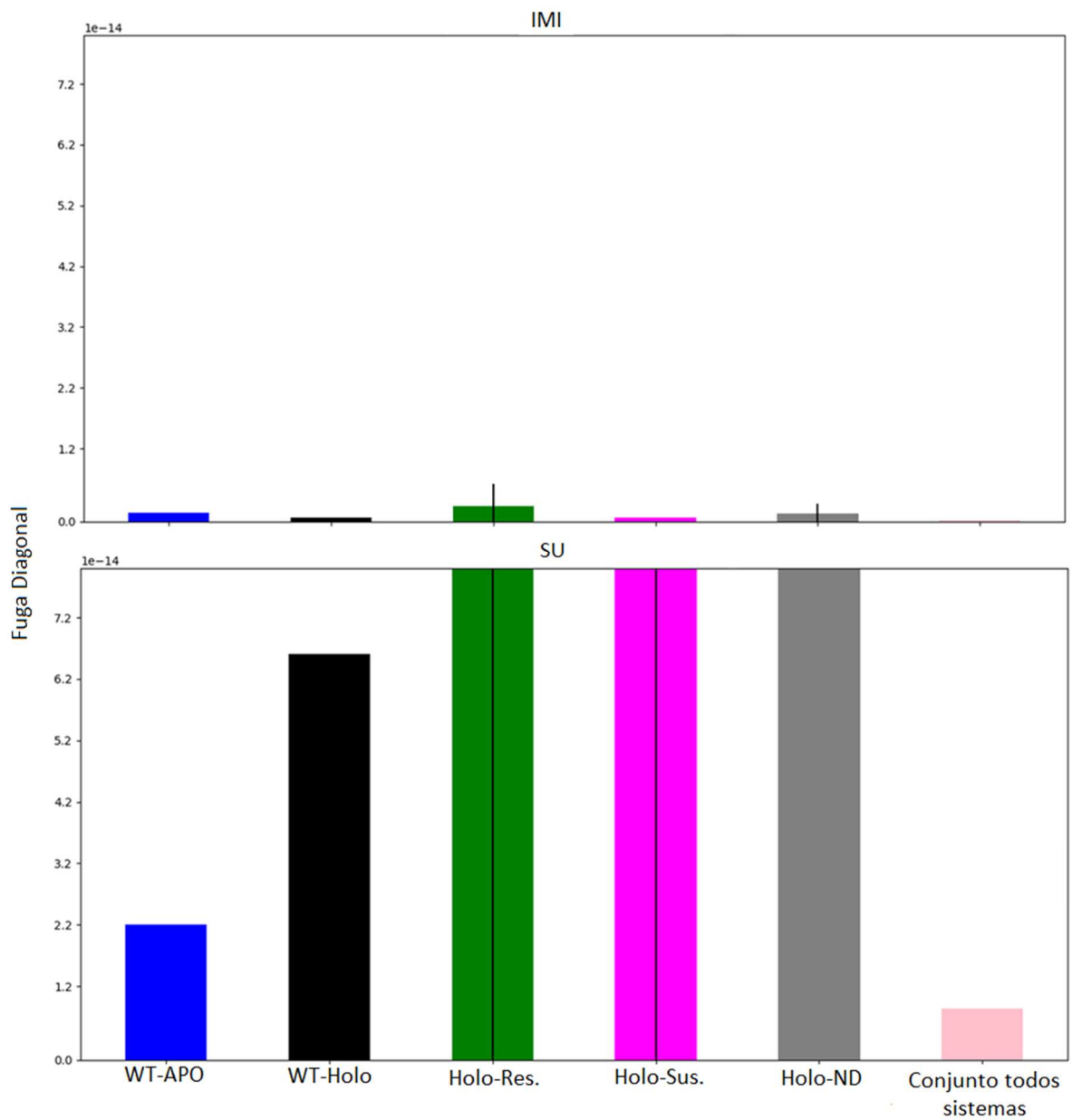


Figura 19. Gráficos de barras para valores de fuga diagonal calculados com resíduos descritores para todos os sistemas.

Fonte: próprio autor.

4.5 Análise de PCA

4.5.1 PyReweighting

Análises de mobilidade dos sistemas (considerando apenas os resíduos descritores) foram feitas através da reconstrução dos perfis de energia livre para cada sistema (WT/APO, WT/holos e holos/mutantes) com o kit de ferramentas PyReweighting2D. Para sistemas com IMI a separação entre proteínas inibidas (susceptíveis e holo) e não inibidas (resistentes e apo) se deu principalmente em projeções de PC1x PC2 que são os movimentos de maior amplitude e intrínsecos da proteína, corroborando os achados de sobreposição entre PCA e modos normais. Um padrão de distribuição entre inibidos e não inibidos mostra que inibidos permanecem na centralidade das PC1xPC2, enquanto que os não inibidos estão dispostos nas extremidades dessas PCs (**Figura 20**). Já para SU a separação entre proteínas inibidas e não inibidas também é possível utilizando as PCs de maior amplitude PC1 e PC2, mas também uma PC mais interna, PC3 (**Figura 21**), corroborando também os achados de sobreposição entre PCA e modos normais, que por sua vez demonstram que os movimentos de SU são menos abrangentes que os de IMI.

Uma segunda análise foi realizada utilizando para a reconstrução dos perfis de energia livre e PCA, o conjunto de trajetórias dos sistemas ligados a cada um dos herbicidas. Essa segunda análise levou em consideração: i) apenas os CAs dos resíduos descritores, ii) os CAs de todos os resíduos. Ainda para cada um dos dois casos fez-se as análises projetando os valores de PC do cluster mais populoso de cada sistema e também os valores de pc das projeções médias.

O mesmo padrão citado acima para IMI e SU foi observado nessa segunda análise de reconstrução dos perfis de energia. A melhor visualização dessa separação foi notada analisando os gráficos construídos com a projeção do cluster mais populoso de cada sistema (comparando-se com a projeção média) (**Figura 22** e **Figura 23**). Proteínas inibidas tendem a ser projetadas em centralidades de PC1xPC2 e PC1xPC3, enquanto que as não inibidas tendem a se afastar dessas centralidades ou em PC1xPC2 ou em PC1xPC3 (maior visibilidade em SU). Os sistemas com SU se mostraram mais complexos e de difícil separação entre resistentes e susceptíveis, sendo necessários mais estudos para entender a natureza dessa separação.

Além da interpretação dos padrões de distribuição de inibidos e não inibidos, pode-se notar que a seleção dos resíduos descritores de movimento pelo aprendizado de máquina se mostra eficaz.

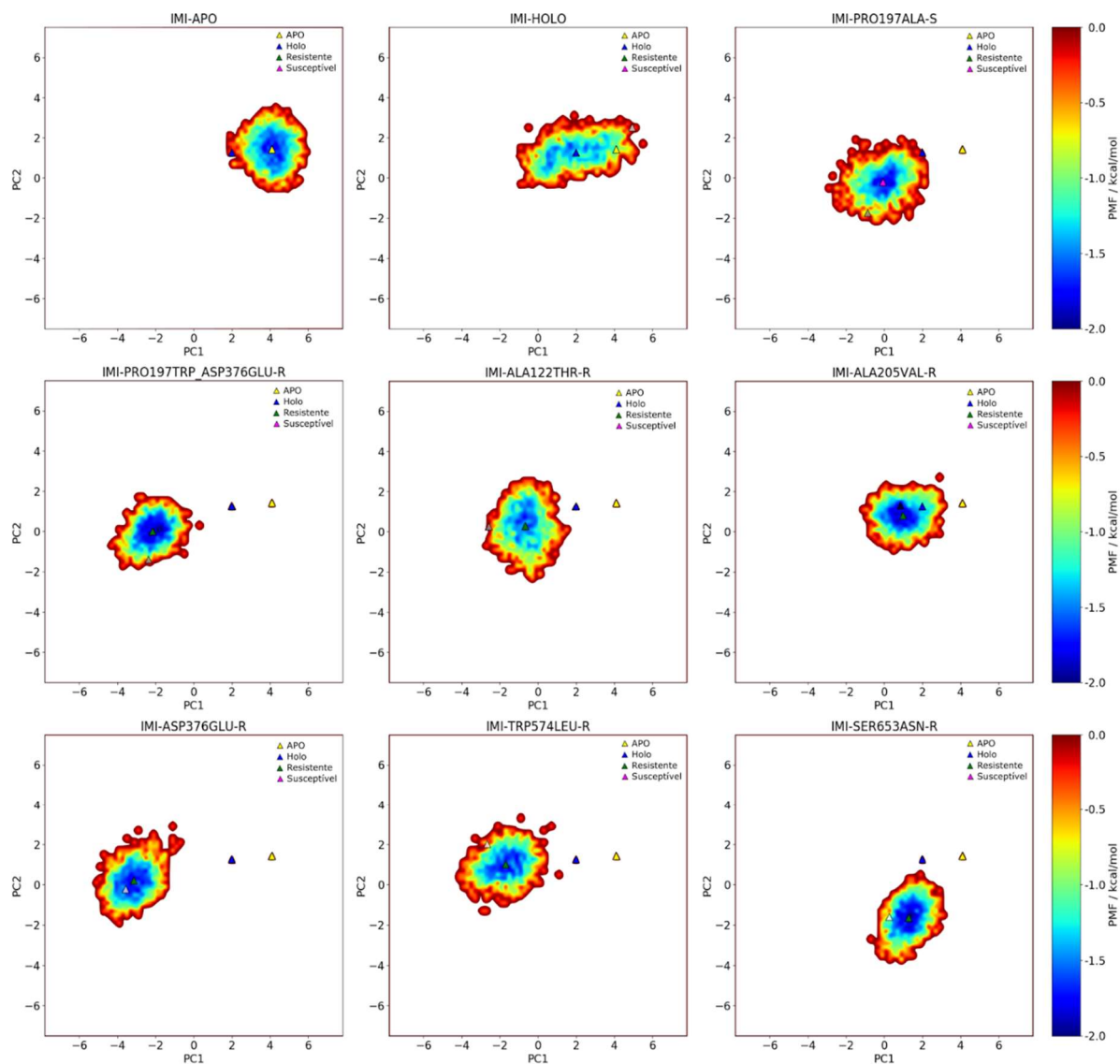


Figura 20. Perfis de energia livre para cada sistema ligado a IMI e para sistema APO. Os triângulos ao longo dos gráficos representam as projeções médias de cada sistema nas PCs, sendo amarelos - resistentes, preto - holo, branco - apo e magenta - susceptíveis.

Fonte: próprio autor.

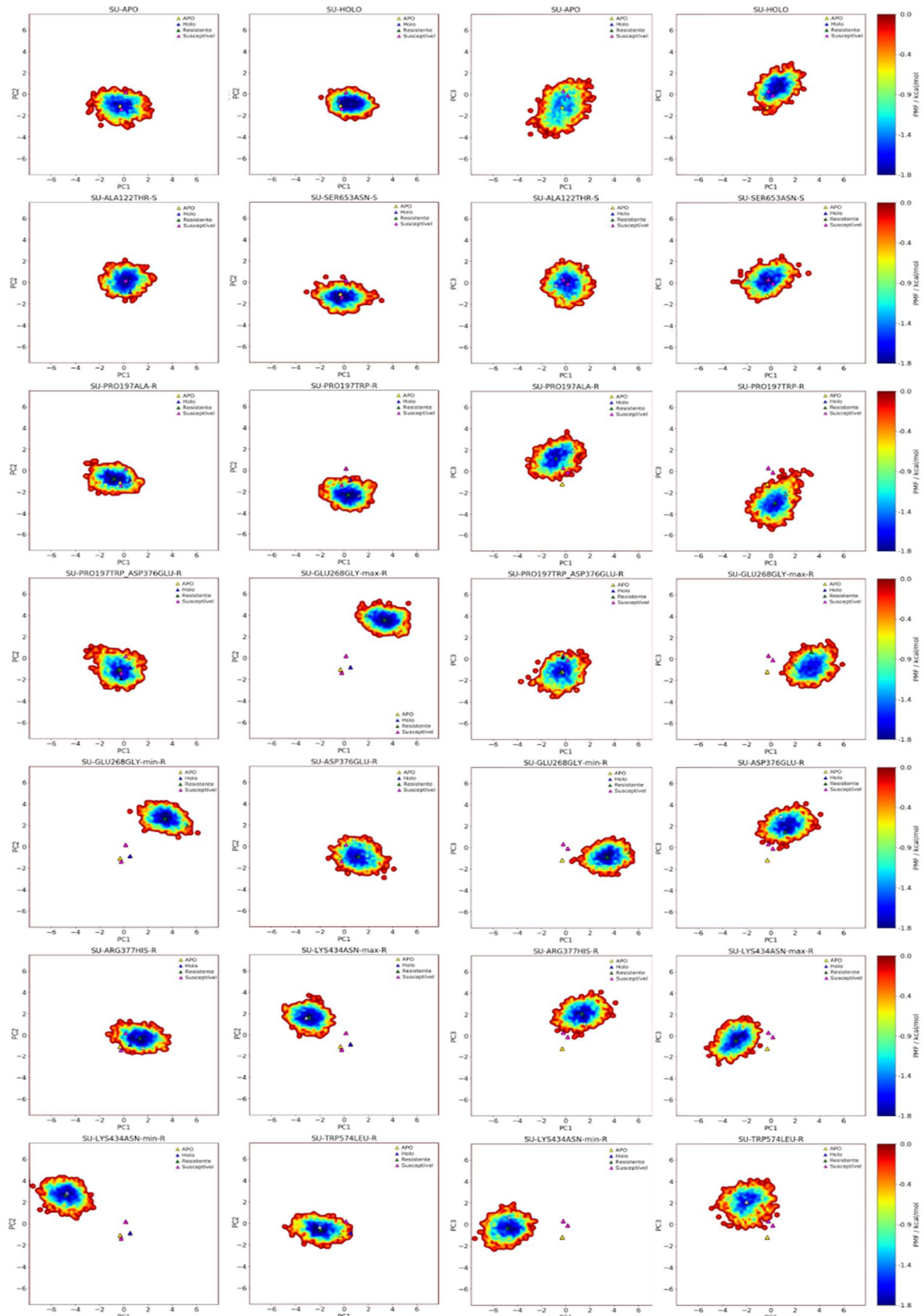


Figura 21. Perfis de energia livre para cada sistema ligado a SU e para sistema APO. Os triângulos ao longo dos gráficos representam as projeções médias de cada sistema ao longo das PCs, sendo amarelos - resistentes, preto - holo, branco - apo e magenta - susceptíveis.

Fonte: próprio autor.

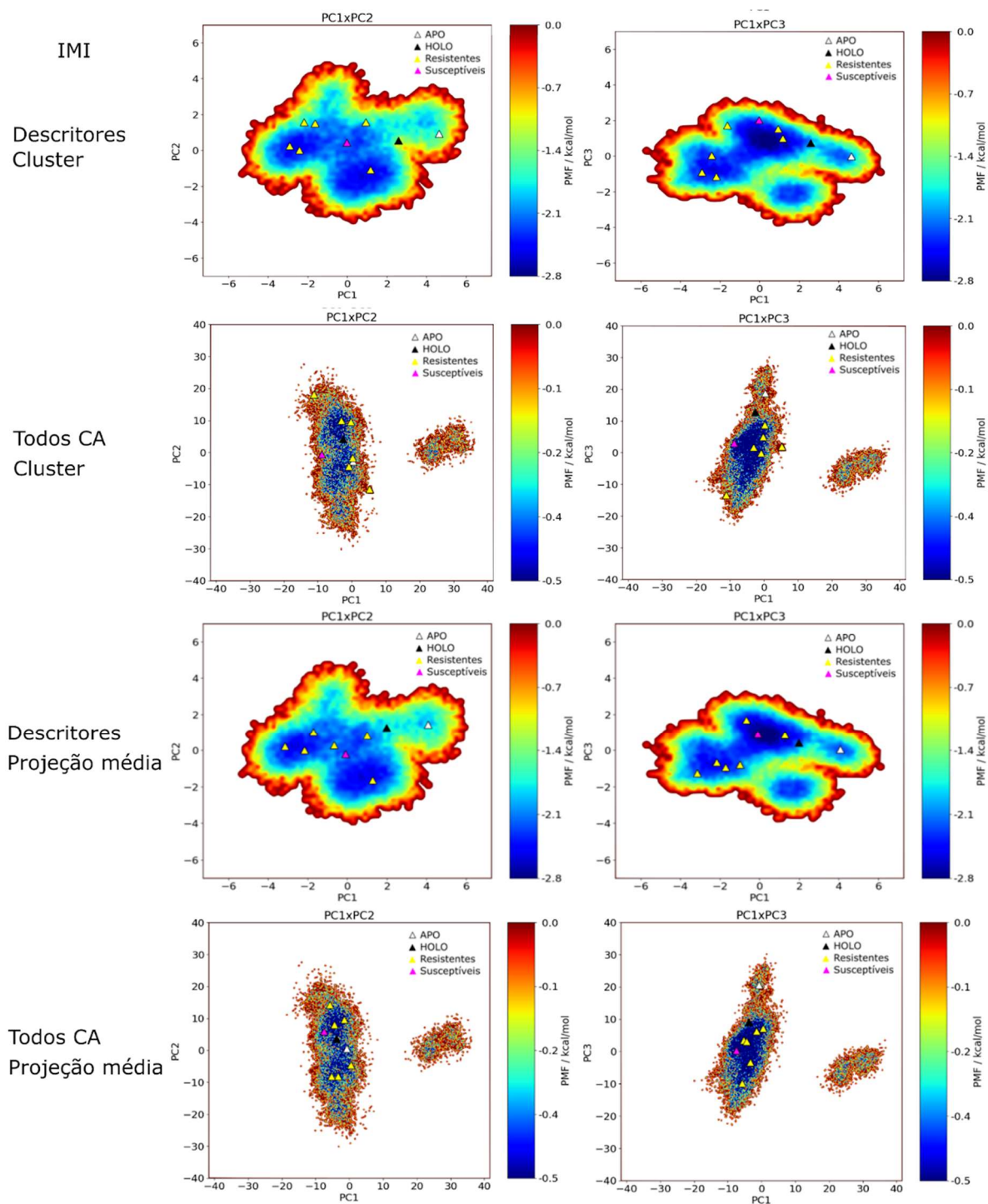


Figura 22. Perfis de energia livre para o conjunto de trajetórias IMI, levando em consideração os CA de todos os resíduos e os CA de resíduos descritores. Os triângulos ao longo do gráfico representam as projeções médias de cada sistema e cluster mais populoso de cada sistema ao longo das PCs. Triângulos amarelos - resistentes, preto - holo, branco - apo e magenta - susceptíveis.

Fonte: próprio autor.

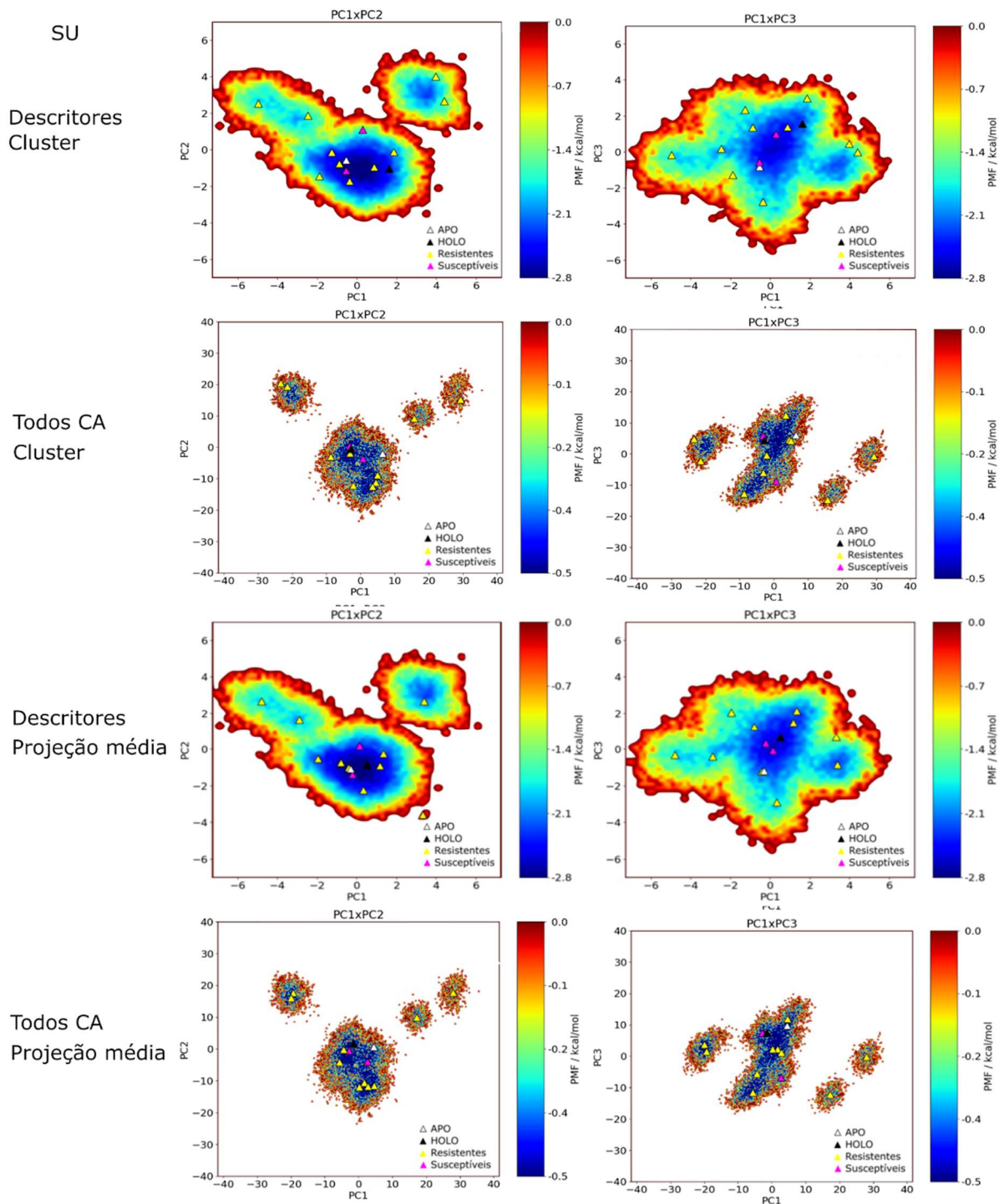


Figura 23. Perfis de energia livre para o conjunto de trajetórias SU, levando em consideração os CA de todos os resíduos e os CA de resíduos descritores. Os triângulos ao longo do gráfico representam as projeções médias de cada sistema e cluster mais populoso de cada sistema ao longo das PCs. Triângulos amarelos - resistentes, preto - holo, branco - apo e magenta – susceptíveis.

Fonte: próprio autor.

4.5.2 Movimentos das componentes principais

Os movimentos amostrados durante a MD para os conjuntos de trajetórias com ligante IMI ou com SU foram recuperados pela análise de PCA levando em consideração apenas os resíduos descritores respectivos de sistemas com IMI ou SU.

Como visto nas análises de perfis de energia livre, os sistemas inibidos ou não inibidos pelos ligantes podem ser separados pelos movimentos principais amostrados durante a MD. No caso do ligante IMI, os sistemas foram melhor separados já nos dois primeiros PCs, já para os sistemas com o ligante SU, um PC menos abrangente foi necessário para uma melhor separação.

A **Figura 24** e **Figura 25** mostram o direcionamento dos movimentos adotados pelos resíduos descritores. Esses resíduos, mesmo mais distantes do sítio ativo no caso de IMI, estão se deslocando de forma a induzir distorções no sítio ativo e próximos aos cofatores. Tanto para SU quanto para IMI, as interações e forma de ligação aos herbicidas sofrem com a forma de movimento dos descritores, promovendo o efeito de inibição ou não inibição em ambos os casos. Para SU, esses movimentos parecem estar menos associados aos movimentos intrínsecos da proteína e mais relacionados a uma mudança desses movimentos. Para o IMI, esses movimentos parecem seguir os movimentos intrínsecos da proteína, se diferenciando em intensidade.

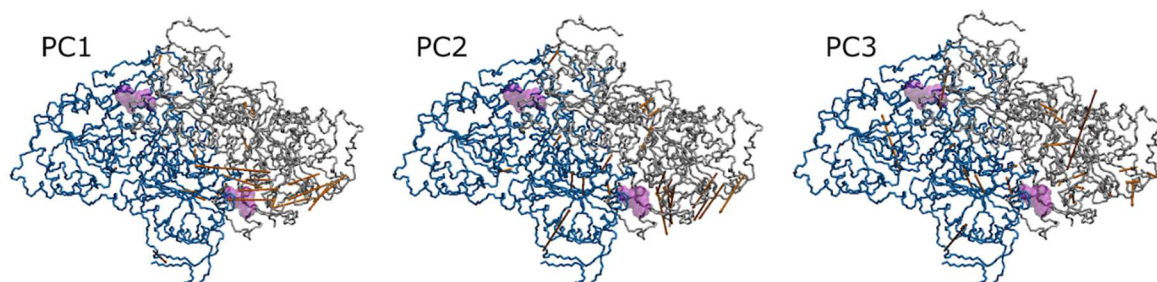


Figura 24. Movimentos adotados por resíduos descritores de sistemas ligados a IMI nas componentes principais PC1, PC2, PC3. Monômeros coloridos de azul escuro e cinza escuro, esferas roxas transparentes representam um herbicida ligado ao sítio. Setas alaranjadas representam a direção do movimento previsto para os resíduos descritores.

Fonte: próprio autor.

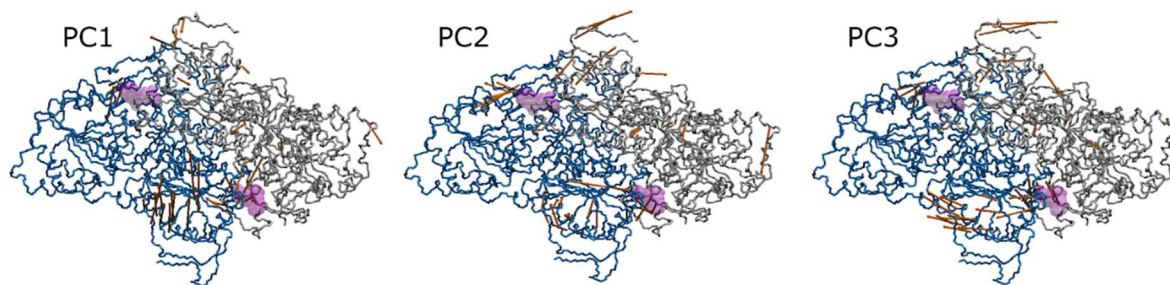


Figura 25. Movimentos adotados por resíduos descritores de sistemas ligados a SU nas componentes principais PC1, PC2, PC3. Monômeros coloridos de azul escuro e cinza escuro, esferas roxas transparentes representam um herbicida ligado ao sítio. Setas alaranjadas representam a direção do movimento previsto para os resíduos descritores.

Fonte: próprio autor.

4.6 Análises de afinidade pelo ligante

4.6.1 Análise de metadinâmica

As análises de metadinâmica de arranque de ligante foram realizadas a fim de se obter valores de $\Delta\Delta G$ (ΔG do sistema holo/mutante - ΔG de sistemas WT/holo) que representem a intensidade de ligação do herbicida à proteína. Uma fraca interação com o ligante é atribuída a valores de $\Delta\Delta G$ positivos e assim uma não inibição da enzima. Uma forte interação com o ligante é atribuída a valores de $\Delta\Delta G$ negativos e assim uma inibição da enzima.

Para sistemas com IMI, a acurácia da técnica, ou seja, acertos em relação à resistência ou susceptibilidade descritas na literatura e valores de $\Delta\Delta G$ positivos ou negativos foi de 57% (**Figura 26**). Esse baixo valor de acurácia pode ser devido ao número significativamente menor de mutantes experimentalmente descritos na literatura para esse ligante.

Para sistemas com SU, a acurácia da técnica é de 67% (**Figura 27**). Essa melhora para a técnica aplicada sobre SU pode ser devida a um maior número de mutantes experimentalmente descritos na literatura para esse ligante.

Para ambos os casos, mesmo com a melhora de acurácia para SU, a acurácia da técnica não foi tão representativa. Um fator importante nessa análise é levar em consideração a natureza mista dos inibidores, sendo esses não totalmente competitivos. Sendo assim, não necessariamente o fenômeno de inibição vai estar relacionado apenas com a afinidade pelo ligante, para algumas mutações pode ser um efeito relacionado à atividade catalítica, devido a perturbações na presença do ligante e assim as deturpações que acontecem nessas regiões mais distantes do sítio podem ter movimentos alostéricos que vão mudar a ação da proteína.

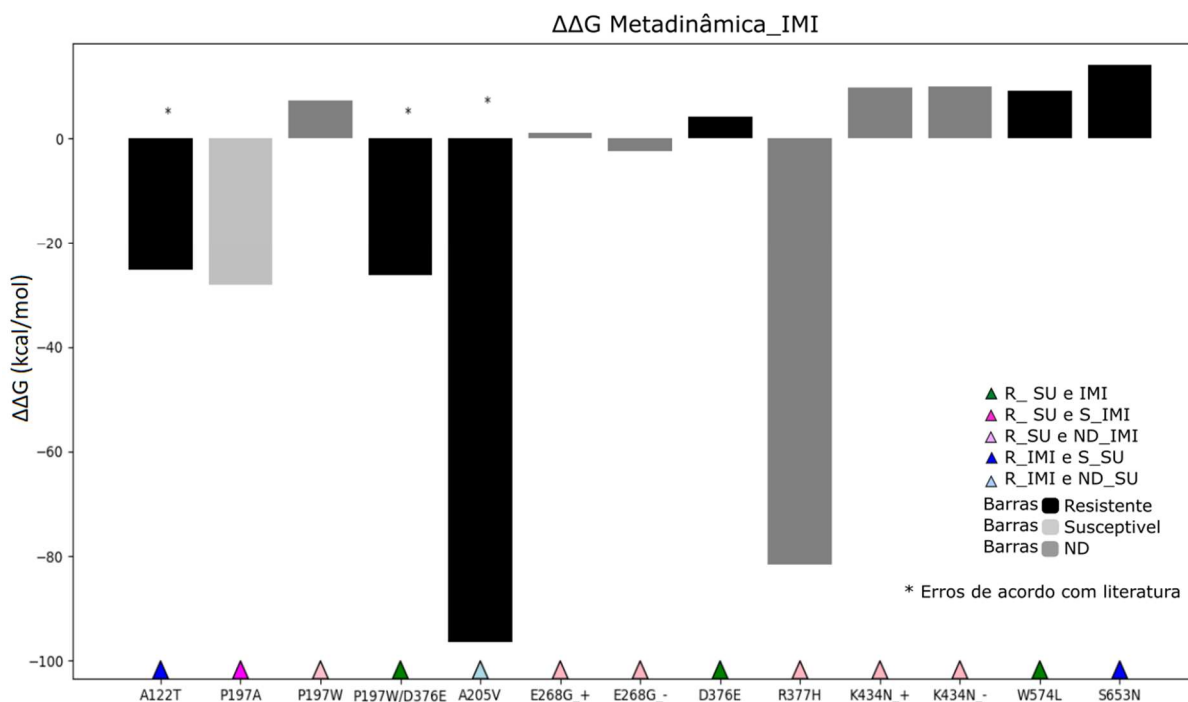


Figura 26. Gráfico de barras para valores de $\Delta\Delta G$ (ΔG do sistema holo/mutantes - ΔG de sistemas WT/holo) para cada sistema com IMI. Barras pretas - sistemas descritos na literatura como resistentes, barras cinza claro - sistemas descritos na literatura como susceptíveis, barras cinza escuro - sistemas não discriminados na literatura. * apontam para sistemas onde a técnica não foi assertiva. Triângulos azuis escuro - mutação de resistência para IMI e não SU, azul claro - mutação de resistência para IMI e não informado para SU, magenta - mutação de resistência para SU e não IMI, rosa claro - mutação de resistência para SU e não informado para IMI e verde - mutação de resistência para IMI e SU.

Fonte: próprio autor.

Sistemas identificados pela técnica como resistentes (valores positivos de $\Delta\Delta G$), parecem sugerir uma maior intensidade dessa resistência relacionada à afinidade com o ligante, em sistemas com SU. Os valores de $\Delta\Delta G$ desses sistemas ficam em torno de vinte a trinta kcal/mol, já para sistemas com IMI os valores não ultrapassam a 10 kcal/mol.

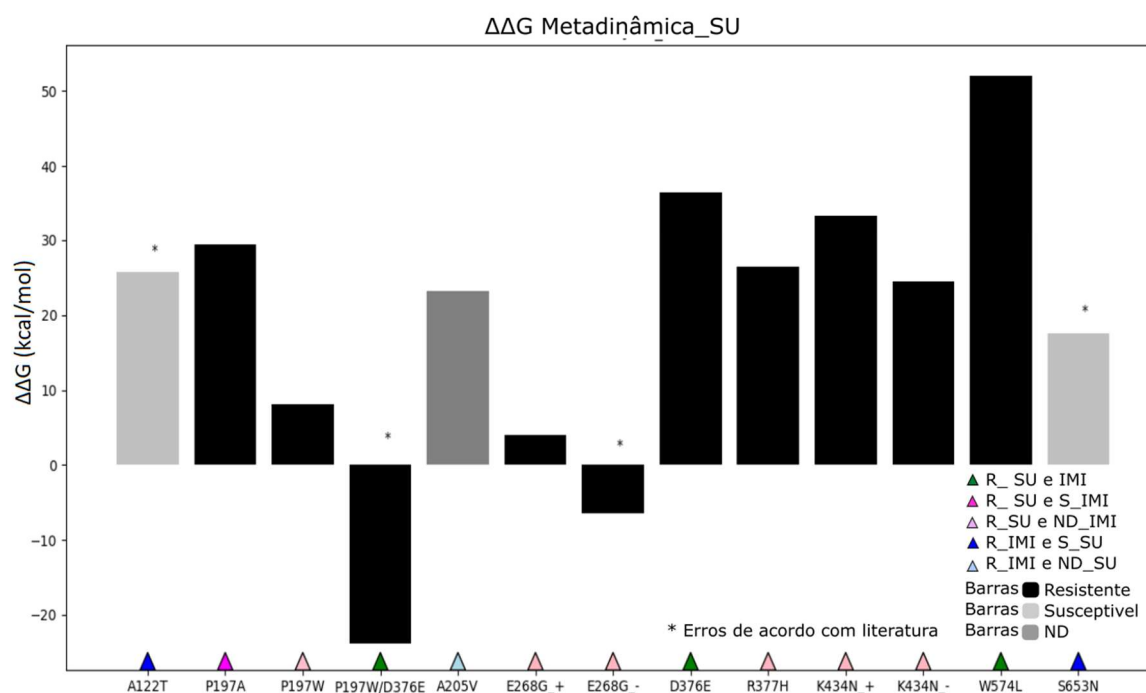


Figura 27. Gráfico de barras para valores de $\Delta\Delta G$ (ΔG do sistema holo/mutantes - ΔG de sistemas WT/holo) para cada sistema com SU. Barras pretas - sistemas descritos na literatura como resistentes, barras cinza claro - sistemas descritos na literatura como susceptíveis, barras cinza escuro - sistemas não discriminados na literatura. * apontam para sistemas onde a técnica não foi assertiva. Triângulos azuis escuro - mutação de resistência para IMI e não SU, azul claro - mutação de resistência para IMI e não informado para SU, magenta - mutação de resistência para SU e não IMI, rosa claro - mutação de resistência para SU e não informado para IMI e verde - mutação de resistência para IMI e SU.

Fonte: próprio autor.

4.6.2 MM/PBSA e Quase-harmônica

Uma segunda abordagem para inferência de energia de ligação dos herbicidas à proteína com e sem mutação foi realizada. Complementando então a técnica de metadinâmica mencionada acima, utilizou-se da técnica de área de superfície de Poisson-Boltzmann da mecânica molecular (MM/PBSA), atrelada à aproximação quase-harmônica para estimativa de entropia. Isso possibilitou a comparação entre sistemas WT/holos e holos/mutantes.

A afinidade pelo ligante pôde ser inferida por diferentes abordagens: i) $\Delta\Delta G$ do complexo ternário (dímero e dois ligantes), ii) $\Delta(\Delta G$ de cooperatividade) e iii) $\Delta\Delta G$ de associação do segundo ligante (**Figura 28** e **Figura 29**). Todos os seguintes dados de $\Delta\Delta G$ foram construídos subtraindo os valores encontrados para as enzimas holo/mutantes dos valores das enzimas WT/holos, a fim de obter uma comparação entre as mesmas ($\Delta\Delta G = \Delta G$ holo/mutante - ΔG WT/holo).

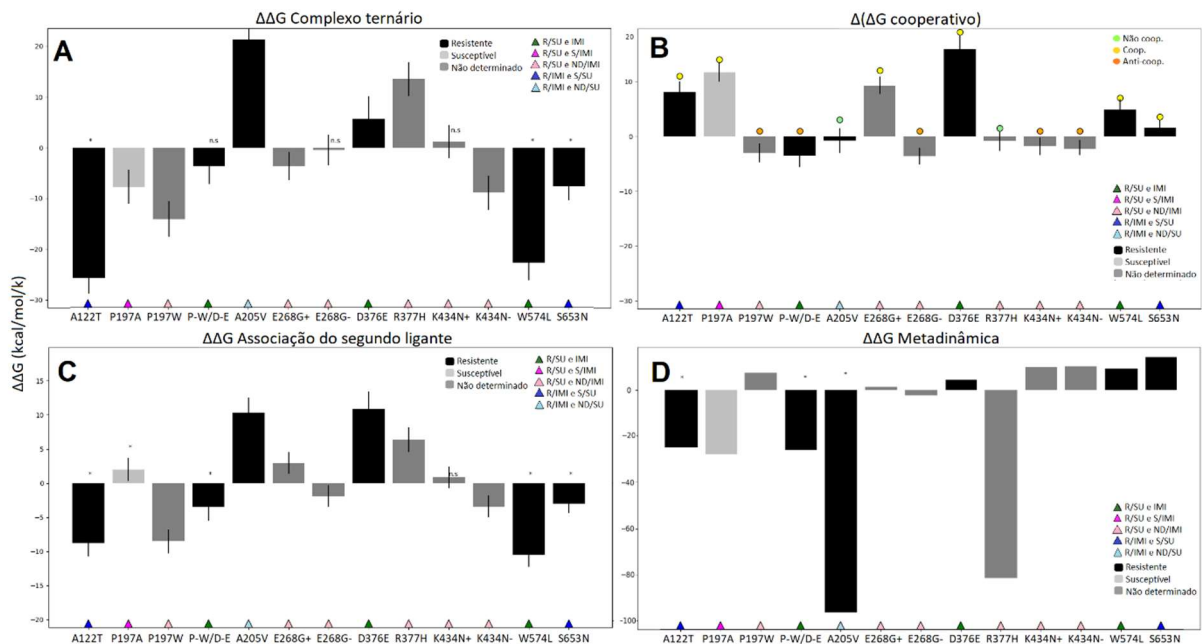


Figura 28. Gráfico de barras para valores de (A) $\Delta\Delta G$ do complexo ternário, (B) $\Delta(\Delta G$ de cooperatividade), (C) $\Delta\Delta G$ de associação do segundo ligante e (D) $\Delta\Delta G$ da metadinâmica, para sistemas ligados a IMI. Barras pretas - sistemas descritos na literatura como resistentes, barras cinza-claro - sistemas descritos na literatura como susceptíveis, barras cinza escuro - sistemas não discriminados na literatura. * apontam para sistemas onde a técnica não foi assertiva. Triângulos azuis escuro - mutação de resistência para IMI e não SU, azul claro - mutação de resistência para IMI e não informado para SU, magenta - mutação de resistência para SU e não IMI, rosa claro - mutação de resistência para SU e não informado para IMI e verde - mutação de resistência para IMI e SU. Círculos verdes $\Delta\Delta G$ não cooperativo, círculos laranja $\Delta\Delta G$ anticooperativo e círculos amarelos $\Delta\Delta G$ cooperativo. Fonte: próprio autor.

Em consonância com os resultados de afinidade calculados através da técnica de metadinâmica, existem alguns sistemas já identificados na literatura como resistentes ou susceptíveis que não apresentam resultados assertivos em relação ao $\Delta\Delta G$ do complexo ternário e $\Delta\Delta G$ de associação do segundo ligante. Mais uma vez pode-se atrelar ao fato de que os inibidores são mistos e possivelmente a resistência ou susceptibilidade pode estar atrelada a movimentos na estrutura que podem não estar diretamente relacionados à afinidade pelo ligante.

As análises de $\Delta\Delta G$ do complexo ternário mostram o quanto o sistema em dímero tem de afinidade por dois ligantes (um em cada sítio ativo), independentemente do caminho e da ordem de ligação dos mesmos. Avaliando os resultados (Figura 28(A) e Figura 29(A)) sistemas ligados a SU possuem uma melhor descrição de resistentes do que sistemas ligados a IMI. Barras pretas (resistentes) são mais assertivas (menor afinidade de ligação) de acordo com a literatura para SU. Demonstrando que o fenômeno de resistência pode estar mais intrinsecamente relacionado à afinidade pelo ligante no caso de SU e menos no caso de IMI.

Os gráficos de associação ao segundo ligante (Figura 28(C) e Figura 29(C)) permitem avaliar a probabilidade de entrada de um segundo ligante após um dos sítios já estar conectado

a um ligante. Valores de $\Delta\Delta G$ mais negativos mostram que existe uma facilidade maior de ligação, valores mais positivos, uma dificuldade maior para a ligação, ou seja, quais os valores de energia gastos durante os processos. Existe uma maior dificuldade para associação do segundo ligante para sistemas com SU, sendo essa dificuldade notada em sua maioria em sistemas resistentes, o que dificultaria a interação com o herbicida no sítio. Já para sistemas com IMI esse padrão não pode ser notado, sistemas resistentes não apresentam o processo de associação dificultado.

As análises de cooperatividade (**Figura 28(B)** e **Figura 29(B)**) relatam o quanto a interação do primeiro ligante no primeiro sítio pode alterar a proteína, e assim tornar mais aceitável ou não a entrada do segundo ligante no segundo sítio, comparada a quando o primeiro sítio ainda não estava ocupado. Nesta análise se busca avaliar a facilidade de ligação de ambos os ligantes, desde que algum consiga interagir, o que a difere da associação do segundo ligante.

Observa-se que existe uma maior cooperatividade em sistemas ligados a SU, valores mais positivos de $\Delta\Delta G$ indicam que existe uma cooperação entre a ligação do primeiro ligante ao primeiro sítio e a ligação do segundo ligante ao segundo sítio. Correlacionando com a entrada do segundo ligante que nos mostra que existe uma dificuldade na ligação do segundo ligante, pode-se inferir que os valores de cooperatividade positiva estão indicando que a entrada do primeiro ligante é mais fácil do que a do segundo, mas quando esse primeiro se encontra no sítio, rapidamente o complexo ternário pode ser formado.

Infere-se que para IMI as modificações na estrutura da proteína após a ligação do primeiro herbicida parecem não estar tão associadas aos sítios ativos. Isso pode ser notado através de uma facilitação maior da entrada do segundo ligante (comparado com SU), uma afinidade maior dos ligantes livres pelo dímero ($\Delta\Delta G$ complexo ternário) e também pela cooperatividade que se mostrou tanto positiva quanto negativa em diferentes sistemas. O fenômeno de modificações estruturais mais próximas aos sítios, após a ligação do herbicida, parece estar mais associado ao ligante SU.

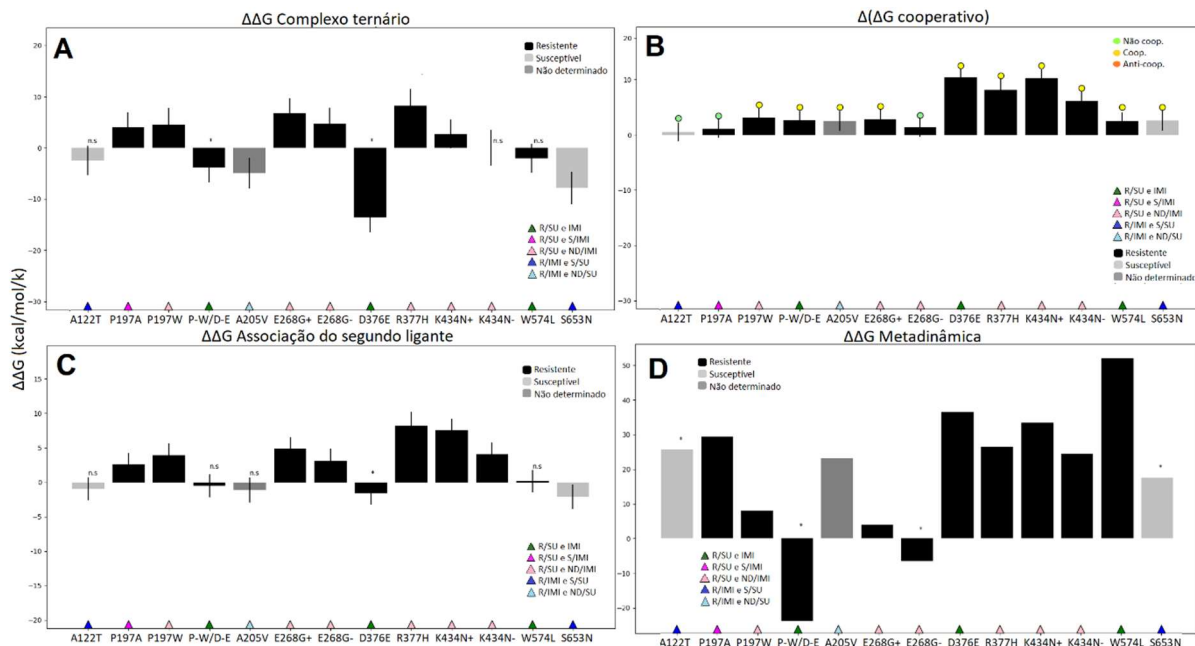


Figura 29. Gráfico de barras para valores de (A) $\Delta\Delta G$ do complexo ternário, (B) $\Delta(\Delta G)$ de cooperatividade, (C) $\Delta\Delta G$ de associação do segundo ligante e (D) $\Delta\Delta G$ da metadinâmica, para sistemas ligados a SU. Barras pretas - sistemas descritos na literatura como resistentes, barras cinza-claro - sistemas descritos na literatura como susceptíveis, barras cinza escuro - sistemas não discriminados na literatura. * apontam para sistemas onde a técnica não foi assertiva. Triângulos azuis escuro - mutação de resistência para IMI e não SU, azul claro - mutação de resistência para IMI e não informado para SU, magenta - mutação de resistência para SU e não IMI, rosa claro - mutação de resistência para SU e não informado para IMI e verde - mutação de resistência para IMI e SU. Círculos verdes $\Delta\Delta G$ não cooperativo, círculos laranja $\Delta\Delta G$ anticooperativo e círculos amarelos $\Delta\Delta G$ cooperativo. Fonte: próprio autor.

4.7 Análise farmacofórica

Para verificação dos contatos atômicos existentes entre os herbicidas, cofatores e a proteína, identificando contatos importantes para ligação dos mesmos em proteínas resistentes e susceptíveis, o web server nAPOLI foi utilizado. Esse servidor incorpora um conjunto de algoritmos a fim de detectar interações do tipo proteína-ligante. Foram inseridas como entrada para o nAPOLI, conformações encontradas nas dinâmicas de cada sistema WT/holo e holo/mutante, que continham coordenadas de projeção próximas às PCs descritas na análise de PCA realizada com resíduos descritores.

Inúmeros contatos foram retornados para cada um dos sistemas. Visando uma abordagem mais assertiva e mais preditiva, a ocorrência ou não dos contatos retornados pelo nAPOLI para cada sistema foi submetida a seleção de atributos novamente implementada no WEKA, sendo que sistemas com IMI foram inseridos em uma análise e sistemas com SU em outra. Buscou-se por contatos importantes para inclusão ou não do sistema na classe de resistência ou susceptibilidade a IMI e SU. A **Tabela 3** mostra quais os contatos importantes

retornados para sistemas com IMI e em quais sistemas (resistentes ou susceptíveis) os contatos foram identificados. A **Tabela 4** apresenta os resultados de SU.

Pode-se notar que sistemas resistentes ligados a IMI compartilham mais ligações com os cofatores (4 contatos) do que com o herbicida (1 contato). Já no caso de SU os sistemas resistentes compartilham 5 contatos importantes com os cofatores e 3 contatos com o herbicida.

Um classificador MLP foi construído para IMI e outro para SU levando em consideração apenas os contatos retornados pela seleção de atributos. Para SU, dois sistemas foram classificados incorretamente (14,30%) e para IMI todos os sistemas foram classificados corretamente.

Tabela 3. Contatos importantes para a classificação em resistentes e susceptíveis retornados pela seleção de atributos para sistemas com IMI.

Contatos	Sistemas resistentes	Sistemas Susceptíveis
IMI:OC'_ARG:199_Attractive		WT/Holo
IMI:C10_ALA:122_Hydrophobic		WT/Holo
IMI: OC'_ ARG:377_Hydrogen-bond		PRO197ALA
IMI:O6_LYS:256_Hydrogen-bond	ALA205VAL SER653ASN TRP574LEU ASP376GLU	
FAD:C8_PHE:206_Hydrophobic	TRP574LEU	PRO197ALA
FAD:O3B_ARG:246_Hydrogen-bond		WT/Holo
FAD:O5B_ARG:373_Hydrogen-bond	TRP574LEU ASP376GLU ALA205VAL	
FAD:N3A_ILE:396_Hydrogen-bond	ALA205VAL ASP376GLU	PRO197ALA WT/Holo
FAD:O1A_ASP185_Repulsive		WT/Holo
TDM: O7_GLY:569_Hydrogen-bond	ASP376GLU TRP574LEU	WT/Holo PRO197ALA

Analisando estruturalmente os contatos retornados, nota-se que alguns contatos presentes apenas em sistemas resistentes acarretam em mudança na estrutura próxima ao cofator no caso de IMI (**Figura 30**). O contato IMI:O6_LYS:256_Hydrogen-bond é reconhecido em sistemas resistentes e é ausente em susceptíveis. A presença desse contato parece alterar a ligação de IMI de modo que o ligante fique mais próximo desse resíduo. Em sistemas susceptíveis o contato IMI: OC'_ ARG:377_Hydrogen-bond parece deixar o FAD um pouco mais livre. Em resistentes a ausência desse contato acarreta na modificação da posição de uma segunda ARG, a ARG373. O contato FAD:O5B_ARG:373_Hydrogen-bond parece estabilizar

o grupamento fosfato presente no FAD. Essa mudança no padrão de contatos é, portanto, propagada até o cofator FAD. Em sistemas resistentes a perda do contato entre a ARG377 e IMI parece deixar mais rígido o anel de flavina presente no FAD que antes era mais livre com o maior distanciamento da ARG377. Essa flavina se localiza próxima à região de encaixe do piruvato no ThDP durante a catálise, o que pode justificar o efeito de resistência mesmo com a presença do inibidor no sítio ativo.

Tabela 4. Contatos importantes para a classificação em resistentes e susceptíveis retornados pela seleção de atributos para sistemas com SU.

Contatos	Sistemas resistentes	Sistemas Susceptíveis
FAD:O1A_ ASP:185_Repulsive	GLU268GLY PRO197ALA TRP574LEU LYS434ASN ARG377HIS ASP376GLU	
FAD:O1P_ ASP:185_Repulsive		SER653ASN
FAD:O5'_ ASP:185_Repulsive	PRO197ALA TRP574LEU LYS434ASN ASP376GLU PRO197TRP	
FAD:O2A_ ARG:373_Hydrogen-bond	ASP376GLU ARG377HIS GLU268GLY LYS434ASN PRO197ALA TRP574LEU	
FAD:O2P_ ARG:373_Hydrogen-bond		SER653ASN
FAD:O5'_ THR:331_Hydrogen-bond		WT/Holo ALA122THR
SU:C15_ TRP:574_Aromatic-stacking	PRO197TRP LYS434ASN PRO197ALA	
SU:N16_ TRP:574_Aromatic-stacking	LYS434ASN ARG377HIS PRO197ALA	
SU:C14_ TRP:574_Hydrophobic	GLU268GLY ARG377HIS	
TDM:N1'_ TYR:118_Aromatic-stacking	TRP574LEU ARG377HIS LYS434ASN GLU268GLY	

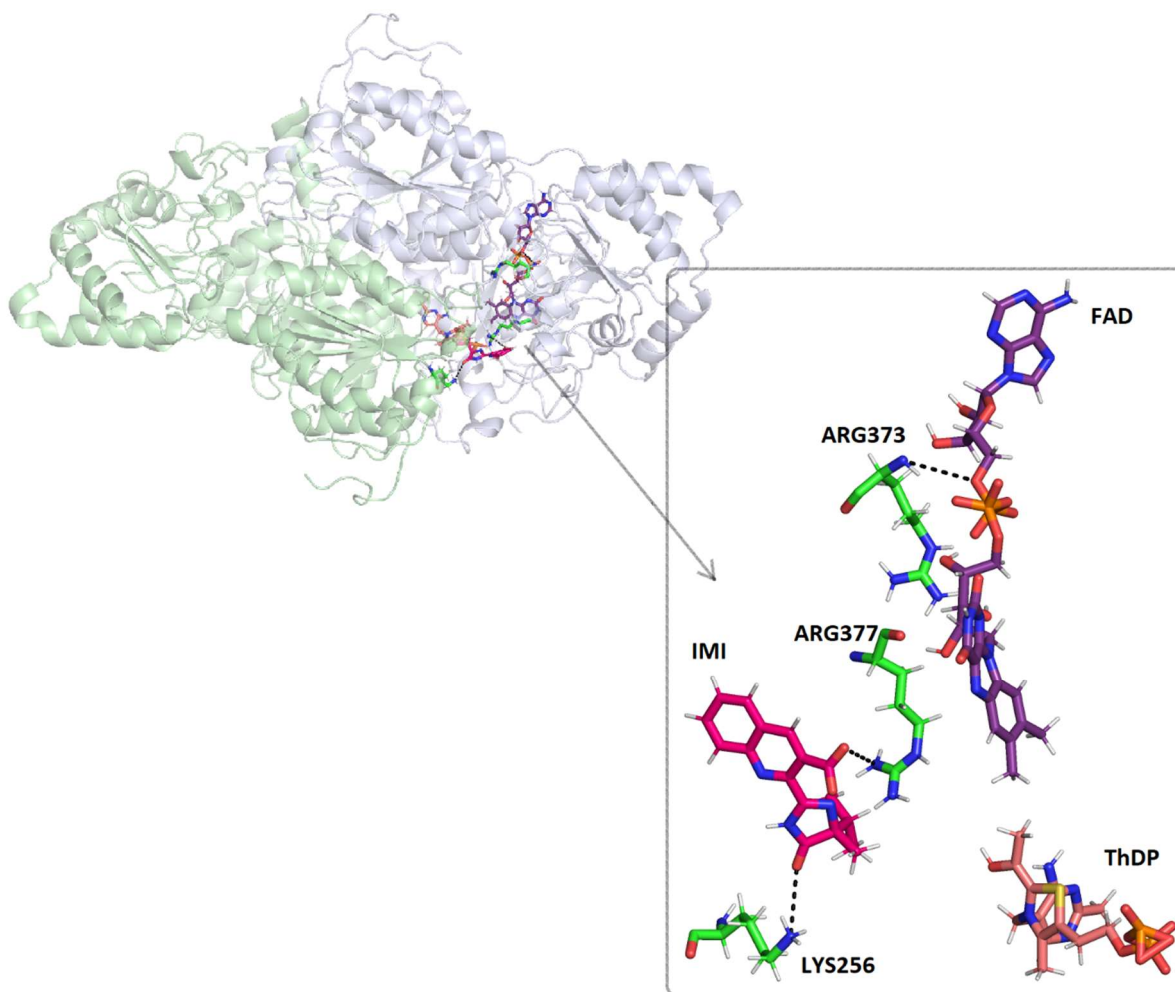


Figura 30. Contatos retornados após a seleção de atributos para IMI na estrutura da proteína WT/holo. Carbonos de IMI em magenta, de FAD em roxo, de ThDP em salmão e de resíduos da proteína em verde. Nitrogênio em azul, hidrogênios em branco, oxigênios em vermelho, fósforo em laranja e enxofre em amarelo. Linhas pontilhadas representam a distância entre os atômicos que fazem contato.
Fonte: próprio autor.

Analisando os contatos retornados para sistemas com SU, nota-se a presença de interações importantes também com os cofatores, incluindo a mesma região da ARG373 que aparece em IMI (**Figura 31**), o que nos influencia a concluir que, a região próxima aos fosfatos presentes no cofator FAD é uma região importante para a atividade da enzima mesmo na presença do inibidor.

Além da interação direta com os cofatores, para SU nota-se dois contatos importantes com o herbicida entre os sistemas resistentes: SU:C15_ TRP:574_Aromatic-stacking SU:N16_ TRP:574_Aromatic-stacking (**Figura 32**). O resíduo TRP574 em sistemas susceptíveis está mais distante do herbicida quando comparado com sistemas resistentes. Essa aproximação parece alterar a região de ligação e assim dificultar a atividade do ligante, ou seja, mesmo quando o ligante se encontra no sítio não consegue cumprir seu papel de inibição da

enzima.

O resíduo TRP574 é relatado na literatura como sendo mutado em ALSs e acarretando na resistência. O aprendizado de máquina em sistemas sem essa mutação parece encontrar como componente de inibição forte a interação direta com o ligante. Já para sistemas mutantes, como TRP574ALA, que não possuem a presença do TRP574, o aprendizado de máquina computa interações com os cofatores FAD e ThDP como sendo mais importantes. Esses achados são consonantes com as análises anteriores. Ambos herbicidas possuem características de inibição mistas que são evidenciadas no aprendizado de máquina em diferentes proporções de acordo com os diferentes ligantes e diferentes mutações.

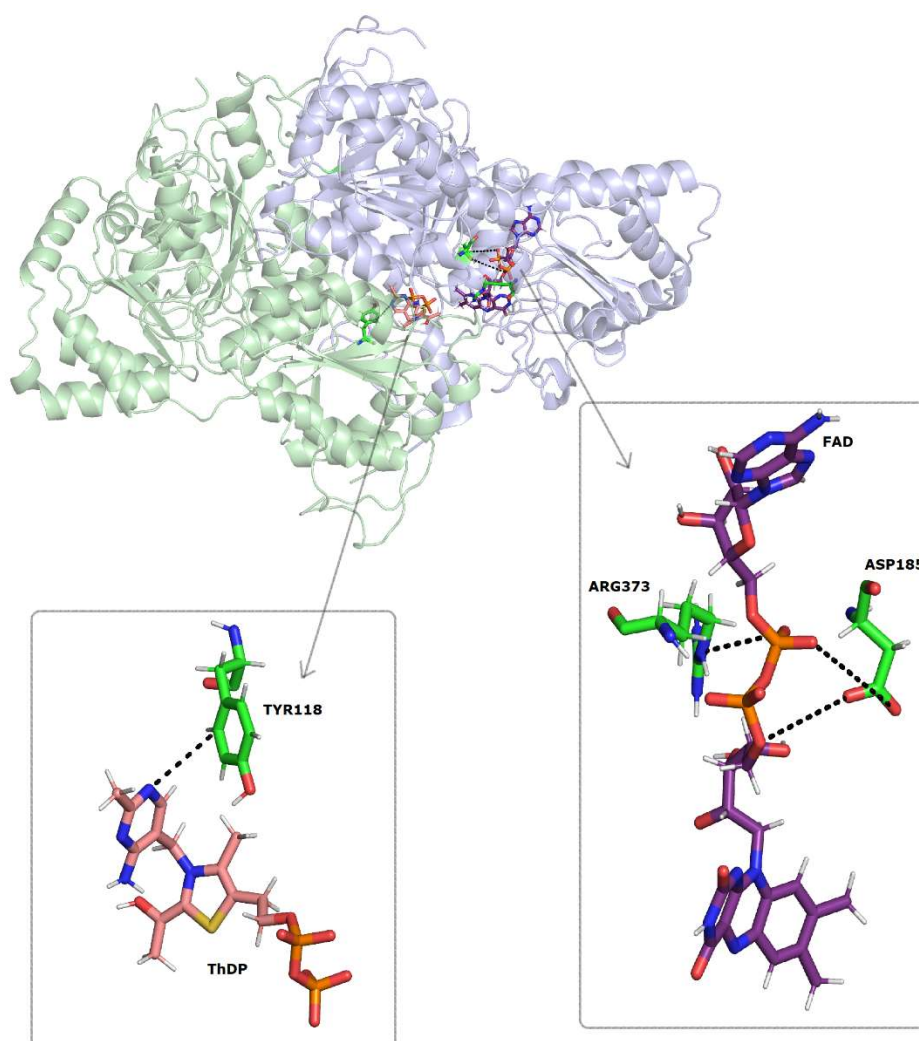


Figura 31. Contatos retornados para SUentre proteína-FAD e proteína_THDP na estrutura da proteína WT/holo e holo/mutantes resistente. Carbonos de SU em magenta, de FAD em roxo, de ThDP em salmão e de resíduos da proteína em verde. Nitrogênio em azul, hidrogênios em branco, oxigênios em vermelho, fosforo em laranja e enxofre em amarelo. Linhas pontilhadas representam a distância entre os atômicos que fazem contato. Distância 4.6 nos contatos de WT/holo e de 3.6 e 3.9 nos contatos de holo/mutante/resistente.

Fonte: próprio autor.

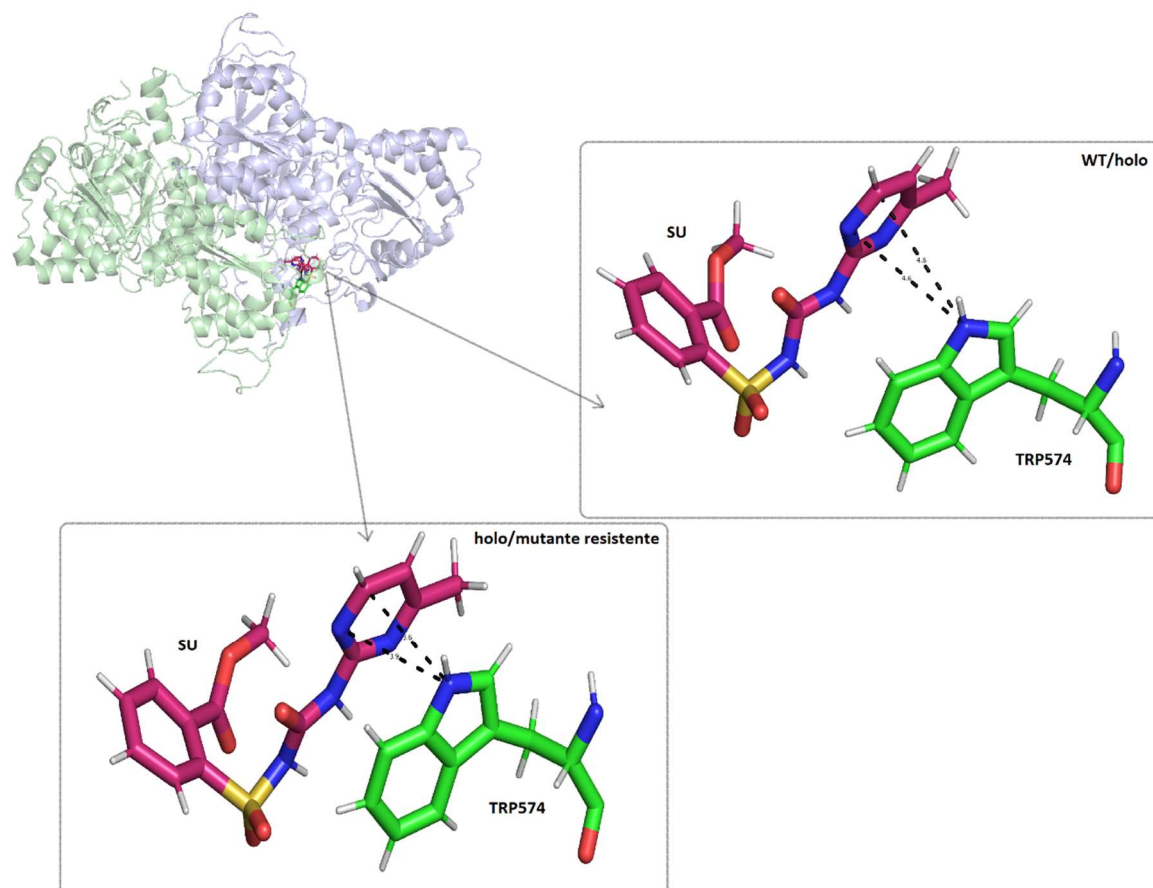


Figura 32. Contatos retornados para SU na estrutura da proteína WT/olo e holo/mutantes resistente. Carbonos de SU em magenta, de FAD em roxo, de ThDP em salmão e de resíduos da proteína em verde. Nitrogênio em azul, hidrogênios em branco, oxigênios em vermelho, fósforo em laranja e enxofre em amarelo. Linhas pontilhadas representam a distância entre os atômicos que fazem contato. Distância 4.6 nos contatos de WT/olo e de 3.6 e 3.9 nos contatos de holo/mutante/resistente.

Fonte: próprio autor.

4.8 Aprendizado de máquina e dados de MD

Um grande volume de dados foi gerado ao final desse estudo, iniciando-se pelos resultados obtidos na separação dos sistemas em PCs (resíduos descritores, todos os resíduos, cluster e projeção média), resultados de afinidade (metadinâmica e MM/PBSA) e finalizando com resultados da análise de contatos entre os herbicidas, cofatores e a proteína.

Como visto anteriormente, o aprendizado de máquina aliado a dados de MD mostrou-se promissor possibilitando análises de grande volume e complexidade de dados de forma mais sistemática e facilitando a interpretação. Visando a junção de vários dados de MD pós

tratamento e possível interpretação dos mesmos, mais uma vez buscou-se pelo aprendizado de máquina.

Em uma busca por padrões em sistemas mutantes resistentes e susceptíveis, prever quais são os dados que melhor separam esses sistemas pode nos informar possíveis próximos caminhos nessa jornada por herbicidas mais eficazes.

Nessa etapa de seleção dos melhores atributos separadores de sistemas com IMI e SU, os conjuntos de dados recuperados nesse trabalho foram inseridos para cada sistema identificado como susceptível ou resistente. Os dados utilizados foram:

- i) valores de PC1 a PC5 da projeção média dos resíduos descritores,
- ii) valores de PC1 a PC5 da projeção média de todos os resíduos,
- iii) valores de PC1 a PC5 da projeção dos clusters dos resíduos descritores,
- iv) valores de PC1 a PC5 da projeção dos clusters de todos os resíduos,
- v) valores de $\Delta(\Delta G \text{ Cooperativo})$,
- vi) valores de $\Delta\Delta G$ do complexo ternário,
- vii) valores de $\Delta\Delta G$ da entrada do segundo ligante,
- viii) valores de $\Delta\Delta G$ da metadinâmica.

A **Tabela 5** apresenta os descritores gerais retornados pela seleção de atributos. Um classificador MLP foi construído para IMI e outro para SU. Para SU um sistema foi classificado incorretamente (7,15 %) e para IMI dois sistemas (22,22%).

Para analisar a eficácia da junção dos contatos retornados pela seleção de atributos juntamente com os atributos retornados nessa segunda análise dos descritores gerais, construiu-se um outro MLP. Para SU dois sistemas foram classificados incorretamente (14,30 %) e para IMI nenhum sistema. A **Tabela 6** e **Tabela 7** mostram o desempenho desses classificadores.

Tabela 5. Descritores gerais retornados pela seleção de atributos.

Seleção de atributos IMI	Seleção de atributos SU
PC1-Projeção média-Descritores	$\Delta(\Delta G \text{ Cooperativo})$
PC5-Cluster-Descritores	$\Delta\Delta G$ -Entrada do 2 ligante

Tabela 6. Resultados obtidos para o classificador MLP da junção dos melhores contatos com melhores descritores gerais de IMI.

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	res
	1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	sus
Weighted Avg.	1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	

Tabela 7. Resultados obtidos para o classificador MLP da junção dos melhores contatos com melhores descritores gerais de SU.

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1,000	0,222	0,714	1,000	0,833	0,745	1,000	1,000	sus
	0,778	0,000	1,000	0,778	0,875	0,745	1,000	1,000	res
Weighted Avg.	0,857	0,079	0,898	0,857	0,860	0,745	1,000	1,000	

É possível notar que os achados descritos anteriormente estão corroborando essa última análise preditiva, onde para IMI os atributos mais importantes para a classificação, além dos contatos com o ligante e os cofatores, são os movimentos estruturais da proteína, principalmente o de maior amplitude (PC1-Projeção média-Descritores), onde a mudança dessa estrutura pode ser mais notada pelo programa. Os contatos retornados para IMI parecem ser atributos importantes e bons para a classificação em resistentes e susceptíveis.

Pra SU os movimentos estruturais não foram considerados de extrema importância para a permanência ou não na classe de resistentes e susceptíveis e sim a interação direta com o ligante refletida na seleção dos atributos relacionados à afinidade com o ligante ($\Delta(\Delta G$ Cooperativo) e $\Delta\Delta G$ -Entrada do 2 ligante. A ligação a esse herbicida pode estar sendo afetada por mudanças estruturais mais localizadas no sítio, o que provoca uma diferença nos contatos e assim visualiza-se os valores de afinidade da interação proteína-ligante como sendo importantes para a classificação em resistentes e susceptíveis. Os classificadores mostram que no caso de SU, os atributos relacionados com a afinidade pelo ligante são melhores separadores das classes.

5. Conclusões

Culturas de plantas são continuamente expostas a herbicidas com o intuito de protegê-las de plantas invasoras. A exposição excessiva a esses químicos, por um longo período de tempo, acarretou no favorecimento de plantas daninhas com o biótipo de resistência aos químicos atualmente aplicados. O desenvolvimento de culturas de plantas resistentes aos herbicidas atuais (contribuindo para que apenas as plantas daninhas sejam atingidas), e o desenvolvimento de novos herbicidas se tornaram extremamente necessários. Para resolver essa problemática, técnicas de bioinformática estrutural atreladas ao aprendizado de máquina foram aqui apresentadas como uma forma mais rápida e econômica para elucidar a natureza da resistência da enzima ALS a duas classes de herbicidas atualmente mais utilizados em campo.

O uso de técnicas de aprendizado de máquina para seleção de resíduos descritores a partir de flutuações de MD retornou apenas 24 resíduos para sistemas com IMI e 35 para sistemas com SU. Esses foram suficientes para classificação com 100% de acurácia de proteínas com IMI ou SU resistentes e susceptíveis. Descritores de IMI estão distribuídos ao longo de toda superfície na proteína e tendem a se localizar mais em um monômero com menor representatividade nos sítios ativos e descritores de SU estão mais localizados próximos aos sítios, e em regiões mais móveis como a lateral da proteína.

As análises de sobreposição entre modos normais e MD foi consideravelmente mais representativa quando voltada para os resíduos descritores de movimento, demonstrando para sistemas com IMI que existe uma significativa sobreposição entre os modos normais preditos pela estrutura intrínseca da proteína e movimentos principais amostrados em MD para proteína WT/APO, WT/holos e holos/mutantes. Já para sistemas com SU essa sobreposição é menor, demonstrando que os movimentos realizados pela proteína com o ligante SU são diferentes de quando ligada a IMI.

A separação das projeções médias de conformações em condições de inibição e não inibição em PCs de maior amplitude PC1 e PC2 foi suficiente para uma separação das mesmas para sistemas com IMI. Estando as proteínas em condições de inibição em regiões mais centrais dessa projeção e as não inibidas em regiões mais extremas. Já para sistemas com SU para uma separação das projeções médias de conformações em condições de inibição e não inibição foi necessário o uso de movimentos de maior amplitude, PC1 e PC2 e também movimentos menos abrangentes da proteína como o PC3. Essa separação tanto para IMI quanto para SU foi melhor descrita quando a análise se volta para os resíduos descritores.

Os valores de $\Delta\Delta G$ em diferentes métricas é capaz de separar resistentes e susceptíveis. A menor afinidade por SU em diferentes métricas pelos resistentes foi mais representativa quando comparada com IMI. Quando o aprendizado de máquina foi alimentado com projeções de movimento e valores de afinidade com o ligante, houve para SU o retorno de valores de afinidade como sendo mais importantes para classificação em resistentes ou susceptíveis, enquanto que para IMI a classificação se torna mais viável através dos movimentos dos descritores que influenciam grandemente a dinâmica da proteína.

Percebe-se que os contatos com o ligante e o cofatores se modificam em qualquer mutação, relacionando mais uma vez a identidade dos inibidores, sendo estes mistos, podem estar mais relacionados com o ligante (afinidade pelo ligante) ou com os cofatores (alosterismo). No caso de IMI, pode-se ressaltar as interações próximas aos cofatores FAD e ThDP, onde a estabilização do grupamento fosfato do FAD parece se tornar um descritor importante para resistência. Já em sistemas com SU, o monitoramento das interações diretas do ligante com resíduos do sítio como TRP574 parece descrever mais o fenômeno de resistência o que pode ser utilizado em trabalhos futuros para o monitorar estruturas quanto à resistência ou susceptibilidade.

O mecanismo de inibição está relacionado com a dinâmica da proteína em ambos, IMI e SU. Essa dinâmica, no caso de IMI, tem relação direta com o perfil de interação que os cofatores têm na proteína, sugerindo um perfil de modificação alostérica da atividade da proteína pelo ambiente químico desses cofatores. Contudo, SU tem um forte componente competitivo, em que as mutações de resistência também tendem a reduzir a afinidade e interação com o ligante, reduzindo a afinidade pelo mesmo. Sabendo que os inibidores são mistos e têm diferentes quantidades da inibição competitiva e inibição alostérica, o aprendizado de máquina sugere padrões diferenciados para IMI e SU.

Tomados em conjunto, nossos estudos computacionais, incluindo o aprendizado de máquina, sugerem que o mecanismo de indução de resistência ou susceptibilidade a IMI reside mais no deslocamento do equilíbrio conformacional da proteína que permite uma alteração do ambiente dos cofatores, sendo mais representativo em IMI. Isso parece sugerir uma maior componente de inibição alostérica em IMI que em SU. No entanto o mecanismo de indução de resistência ou susceptibilidade a SU é mais complexo que o de IMI e esse mecanismo tende a ser direcionado para um ajuste conformacional na proteína modulado por distorções locais. Essas distorções refletem na perda de afinidade pelo ligante, e, portanto, a componente de interação direta com o ligante (componente competitiva) é mais forte que a componente de inibição alostérica.

6. Perspectivas

Como perspectivas pretende-se em trabalhos futuros:

- análise de contatos proteína-proteína para as mesmas estruturas das análises farmacofóricas de cada sistema. Essa análise será realizada através do programa VTR ainda a ser publicado pelo grupo de pesquisa da Prof. Dra. Raquel Minardi.
- análise comparativa de conservação de resíduos, mobilidade conformacional e padrões coevolutivos, de estruturas de ALS de vários organismos disponíveis no banco de dados PDB.

Esse trabalho está sendo transferido para um manuscrito para publicação.

7. Referências bibliográficas

- [1] OERKE, E. C. Crop losses to pests. **The Journal of Agricultural Science**, v. 144, p. 31, 2006.
- [2] POWLES, Stephen B.; YU, Qin. Evolution in action: plants resistant to herbicides. **Annual review of plant biology**, v. 61, p. 317-347, 2010.
- [3] VILA-AIUB, Martin M. Fitness of herbicide-resistant weeds: Current knowledge and implications for management. **Plants**, v. 8, n. 11, p. 469, 2019.
- [4] FARTYAL, Dharendra et al. Developing dual herbicide tolerant transgenic rice plants for sustainable weed management. **Scientific reports**, v. 8, n. 1, p. 1-12, 2018.
- [5] DAYAN, Franck E. Current status and future prospects in herbicide discovery. **Plants**, v. 8, n. 9, p. 341, 2019.
- [6] LEON, Ramon G.; DUNNE, Jeffrey C.; GOULD, Fred. The role of population and quantitative genetics and modern sequencing technologies to understand evolved herbicide resistance and weed fitness. **Pest management science**, 2020.
- [7] BECKIE, Hugh J.; ASHWORTH, Michael B.; FLOWER, Ken C. Herbicide resistance management: Recent developments and trends. **Plants**, v. 8, n. 6, p. 161, 2019.
- [8] HAWKINS, Nichola J. et al. The evolutionary origins of pesticide resistance. **Biological Reviews**, v. 94, n. 1, p. 135-155, 2019.
- [9] HEAP, I. The International Survey of Herbicide Resistant Weeds. Online. Internet. Disponível em: <www.weedscience.org>. Acesso em: 23 fev. 2020.
- [10] PALUMBI, Stephen R. Humans as the world's greatest evolutionary force. **Science**, v. 293, n. 5536, p. 1786-1790, 2001.
- [11] PETERS, Bodo; STREK, Harry J. Herbicide discovery in light of rapidly spreading resistance and ever-increasing regulatory hurdles. **Pest management science**, v. 74, n. 10, p. 2211-2215, 2018.
- [12] NEVE, Paul; VILA-AIUB, Martin; ROUX, Fabrice. Evolutionary-thinking in agricultural weed management. **New Phytologist**, v. 184, n. 4, p. 783-793, 2009.
- [13] Heap, I. The International Herbicide-Resistant Weed Database. Disponível em: <www.weedscience.org>. Acesso em: 3 de Dezembro de 2020.
- [14] CHIPMAN, David M.; DUGGLEBY, Ronald G.; TITTMANN, Kai. Mechanisms of acetoxyacid synthases. **Current opinion in chemical biology**, v. 9, n. 5, p. 475-481, 2005.
- [15] CHIPMAN, David; BARAK, Ze'ev; SCHLOSS, John V. Biosynthesis of 2-aceto-2-hydroxy acids: acetolactate synthases and acetoxyacid synthases. **Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology**, v. 1385, n. 2, p. 401-419, 1998.

- [16] MCCOURT, J. A.; DUGGLEBY, R. G. Acetohydroxyacid synthase and its role in the biosynthetic pathway for branched-chain amino acids. **Amino acids**, v. 31, n. 2, p. 173-210, 2006.
- [17] CHIPMAN, David M.; DUGGLEBY, Ronald G.; TITTMANN, Kai. Mechanisms of acetohydroxyacid synthases. **Current opinion in chemical biology**, v. 9, n. 5, p. 475-481, 2005.
- [18] DUGGLEBY, Ronald G.; MCCOURT, Jennifer A.; GUDDAT, Luke W. Structure and mechanism of inhibition of plant acetohydroxyacid synthase. **Plant Physiology and Biochemistry**, v. 46, n. 3, p. 309-324, 2008.
- [19] DUGGLEBY, Ronald G.; MCCOURT, Jennifer A.; GUDDAT, Luke W. Structure and mechanism of inhibition of plant acetohydroxyacid synthase. **Plant Physiology and Biochemistry**, v. 46, n. 3, p. 309-324, 2008.
- [20] TITTMANN, Kai et al. NMR analysis of covalent intermediates in thiamin diphosphate enzymes. **Biochemistry**, v. 42, n. 26, p. 7885-7891, 2003.
- [21] DUGGLEBY, Ronald G.; PANG, Siew Siew. Acetohydroxyacid synthase. **Journal of biochemistry and molecular biology**, v. 33, n. 1, p. 1-36, 2000.
- [22] DUGGLEBY, Ronald G. Domain relationships in thiamine diphosphate-dependent enzymes. **Accounts of chemical research**, v. 39, n. 8, p. 550-557, 2006.
- [23] MCCOURT, Jennifer A. et al. Herbicide-binding sites revealed in the structure of plant acetohydroxyacid synthase. **Proceedings of the National Academy of Sciences**, v. 103, n. 3, p. 569-573, 2006.
- [24] TRANEL, Patrick J.; WRIGHT, Terry R. Resistance of weeds to ALS-inhibiting herbicides: what have we learned?. **Weed Science**, v. 50, n. 6, p. 700-712, 2002.
- [25] YU, Qin; POWLES, Stephen B. Resistance to AHAS inhibitor herbicides: current understanding. **Pest management science**, v. 70, n. 9, p. 1340-1350, 2014.
- [26] GARCIA, Mario Daniel et al. Crystal structure of plant acetohydroxyacid synthase, the target for several commercial herbicides. **The FEBS Journal**, v. 284, n. 13, p. 2037-2051, 2017.
- [27] MALLORY-SMITH, Carol A.; THILL, Donald C.; DIAL, Michael J. Identification of sulfonyleurea herbicide-resistant prickly lettuce (*Lactuca serriola*). **Weed Technology**, p. 163-168, 1990.
- [27] HATAMI, Zahra M. et al. Multiple mechanisms increase levels of resistance in *Rapistrum rugosum* to ALS herbicides. **Frontiers in plant science**, v. 7, p. 169, 2016.
- [28] LI, Yan et al. In silico identification of genetic mutations conferring resistance to acetohydroxyacid synthase inhibitors: A case study of *Kochia scoparia*. **PloS one**, v. 14, n. 5, p. e0216116, 2019.
- [29] HEAP, Ian. Global perspective of herbicide-resistant weeds. **Pest management science**, v. 70, n. 9, p. 1306-1315, 2014.

- [30] TAN, Siyuan et al. Imidazolinone-tolerant crops: history, current status and future. **Pest Management Science: Formerly Pesticide Science**, v. 61, n. 3, p. 246-257, 2005.
- [31] CHANDRA, Nagasuma; ANAND, Praveen; YETURU, Kalidas. Structural bioinformatics: deriving biological insights from protein structures. **Interdisciplinary Sciences: Computational Life Sciences**, v. 2, n. 4, p. 347-366, 2010.
- [32] Protein Data Bank: the single global archive for 3D macromolecular structure data. **Nucleic acids research**, v. 47, n. D1, p. D520-D528, 2019.
- [33] BERMAN, Helen M. The protein data bank: a historical perspective. **Acta Crystallographica Section A: Foundations of Crystallography**, v. 64, n. 1, p. 88-95, 2008.
- [34] BERMAN, Helen M. et al. The Protein Data Bank . **Nucleic Acids Research**, v. 28, p. 235-242, 2000.
- [35] SUÁREZ, Dimas; DÍAZ, Natalia. SARS-CoV-2 main protease: a molecular dynamics study. **Journal of chemical information and modeling**, v. 60, p. 5815 – 5831, 2020.
- [36] FASSIO, Alexandre V. et al. nAPOLI: a graph-based strategy to detect and visualize conserved protein-ligand interactions in large-scale. **IEEE/ACM transactions on computational biology and bioinformatics**, v. 17, n. 4, p. 1317-1328, 2019.
- [37] HALL, Mark Andrew. Correlation-based feature selection for machine learning. 1999.
- [38] Bakan A, Meireles LM, Bahar I *ProDy: Protein Dynamics Inferred from Theory and Experiments* **Bioinformatics** **2011** 27(11):1575-1577.
- [39] DeLano WL (2002) The PyMOL molecular graphics system. <http://www.pymol.org>.
- [40] CASE, David A. et al. The Amber biomolecular simulation programs. **Journal of computational chemistry**, v. 26, n. 16, p. 1668-1688, 2005.
- [41] MILLER III, Bill R. et al. MMPBSA. py: an efficient program for end-state free energy calculations. **Journal of chemical theory and computation**, v. 8, n. 9, p. 3314-3321, 2012.
- [42] Michele Seeber, Marco Cecchini, Francesco Rao, Giovanni Settanni and Amedeo Caflisch; **Wordom: a program for efficient analysis of molecular dynamics simulations**; **Bioinformatics**, 2007, 23(19):2625-2627.
- [43] SINKO, William et al. Population based reweighting of scaled molecular dynamics. **The Journal of Physical Chemistry B**, v. 117, n. 42, p. 12759-12768, 2013.
- [44] MIAO, Yinglong et al. Improved reweighting of accelerated molecular dynamics simulations for free energy calculation. **Journal of Chemical Theory and Computation**, v. 10, n. 7, p. 2677-2689, 2014.
- [45] RUSSELL, Mark H.; SALADINI, John L.; LICHTNER, Frank. Sulfonylurea herbicides. **Pesticide Outlook**, v. 13, n. 4, p. 166-173, 2002.

- [46] WANG, Jian-Guo et al. Crystal structures of two novel sulfonylurea herbicides in complex with *Arabidopsis thaliana* acetohydroxyacid synthase. **The FEBS journal**, v. 276, n. 5, p. 1282-1290, 2009.
- [47] DURNER, Jörg; BÖGER, Peter. Acetolactate synthase from barley (*Hordeum vulgare* L.): purification and partial characterization. **Zeitschrift für Naturforschung C**, v. 43, n. 11-12, p. 850-856, 1988.
- [48] BABCZINSKI, Peter; ZELINSKI, Thomas. Mode of action of herbicidal ALS-inhibitors on acetolactate synthase from green plant cell cultures, yeast, and *Escherichia coli*. **Pesticide science**, v. 31, n. 3, p. 305-323, 1991.
- [49] HILL, M. Craig; PANG, Siew Siew; DUGGLEBY, G. Ronald. Purification of *Escherichia coli* acetohydroxyacid synthase isoenzyme II and reconstitution of active enzyme from its individual pure subunits. **Biochemical Journal**, v. 327, n. 3, p. 891-898, 1997.
- [50] SCHLOSS, John V. Interaction of the herbicide sulfometuron methyl with acetolactate synthase: a slow-binding inhibitor. In: **Flavins and flavoproteins**. de Gruyter, 1984. p. 737-740.
- [51] SCHLOSS, John V. Acetolactate synthase, mechanism of action and its herbicide binding site. **Pesticide Science**, v. 29, n. 3, p. 283-292, 1990.
- [52] AHAN, Tae Woo; KIM, Dae Whang; DO CHOI, Jung. Original Articles; Inhibition of Acetohydroxyacid Synthase by Sulfonylureas and Imidazolinones. **BMB Reports**, v. 25, n. 7, p. 636-641, 1992.
- [53] SHANER, Dale L.; ANDERSON, Paul C.; STIDHAM, Mark A. Imidazolinones: potent inhibitors of acetohydroxyacid synthase. **Plant Physiology**, v. 76, n. 2, p. 545-546, 1984.
- [54] NELSON, D. L.; COX, M. M. **Lehninger Princípios de Bioquímica**. 3 ed. São Paulo: Sarvier, 2002.
- [55] SKJAERVEN, Lars; HOLLUP, Siv M.; REUTER, Nathalie. Normal mode analysis for proteins. **Journal of Molecular Structure: THEOCHEM**, v. 898, n. 1-3, p. 42-48, 2009.
- [56] SKJAERVEN, Lars; MARTINEZ, Aurora; REUTER, Nathalie. Principal component and normal mode analysis of proteins; a quantitative comparison using the GroEL subunit. **Proteins: Structure, Function, and Bioinformatics**, v. 79, n. 1, p. 232-243, 2011.
- [57] LÓPEZ-BLANCO, José Ramón et al. Normal mode analysis techniques in structural biology. **Els**, 2014.
- [58] S SCHULTZ-HEIENBROK, Robert; MAIER, Timm; STRÄTER, Norbert. A Large Hinge Bending Domain Rotation Is Necessary for the Catalytic Function of *Escherichia coli* 5'-Nucleotidase. **Biochemistry**, v. 44, n. 7, p. 2244-2252, 2005.
- [59] HO, Bosco K.; DILL, Ken A. Folding very short peptides using molecular dynamics. **PLoS Comput Biol**, v. 2, n. 4, p. e27, 2006.

- [60] BEVERIDGE, David L. et al. Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA oligonucleotides. I. Research design and results on d (CpG) steps. **Biophysical journal**, v. 87, n. 6, p. 3799-3813, 2004.
- [61] Taylor, N R, and M von Itzstein. 1994. "Molecular Modeling Studies on Ligand Binding to Sialidase from Influenza Virus and the Mechanism of Catalysis." *J. Med. Chem.* 37 (5): 616–24.
- [62] VETTORETTI, Gerolamo et al. Molecular dynamics simulations reveal the mechanisms of allosteric activation of Hsp90 by designed ligands. **Scientific reports**, v. 6, n. 1, p. 1-13, 2016.
- [63] PHILLIPS, James C. et al. Scalable molecular dynamics with NAMD. **Journal of computational chemistry**, v. 26, n. 16, p. 1781-1802, 2005.
- [64] LIMONGELLI, Vittorio; BONOMI, Massimiliano; PARRINELLO, Michele. Funnel metadynamics as accurate binding free-energy method. **Proceedings of the National Academy of Sciences**, v. 110, n. 16, p. 6358-6363, 2013.
- [65] HAENLEIN, Michael; KAPLAN, Andreas. A brief history of artificial intelligence: On the past, present, and future of artificial intelligence. **California management review**, v. 61, n. 4, p. 5-14, 2019.
- [66] HORNIK, Kurt; BUCHTA, Christian; ZEILEIS, Achim. Open-source machine learning: R meets Weka. **Computational Statistics**, v. 24, n. 2, p. 225-232, 2009.
- [67] Bishop CM. *Pattern recognition and machine learning*. New York: Springer; 2006.
- [68] APOLLONI, Bruno et al. (Ed.). **Machine learning and robot perception**. Springer Science & Business Media, 2005.
- [69] AO, Sio-Iong; RIEGER, Burghard B.; AMOUZEGAR, Mahyar (Ed.). **Machine learning and systems engineering**. Springer Science & Business Media, 2010.
- [71] YU, Jun; TAO, Dacheng. **Modern machine learning techniques and their applications in cartoon animation research**. John Wiley & Sons, 2013.
- [73] MALLEY, James D.; MALLEY, Karen G.; PAJEVIC, Sinisa. **Statistical learning for biomedical data**. Cambridge University Press, 2011.
- [74] KHOTANZAD, Alireza; HONG, Yaw Hua. Rotation invariant image recognition using features selected via a systematic method. **Pattern recognition**, v. 23, n. 10, p. 1089-1101, 1990.
- [75] SWETS, Daniel L.; WENG, John J. Efficient content-based image retrieval using automatic feature selection. In: **Proceedings of International Symposium on Computer Vision-ISCV**. IEEE, 1995. p. 85-90.
- [76] VAN LANDEGHEM, Sofie et al. Discriminative and informative features for biomolecular text mining with ensemble feature selection. **Bioinformatics**, v. 26, n. 18, p. i554-i560, 2010
- [77] LIU, Huiqing; LI, Jinyan; WONG, Limsoon. A comparative study on feature selection and classification methods using gene expression profiles and proteomic patterns. **Genome informatics**, v. 13, p. 51-60, 2002.

- [78] ZHENG, Chun-Hou et al. Tumor clustering using nonnegative matrix factorization with gene selection. **IEEE Transactions on Information Technology in Biomedicine**, v. 13, n. 4, p. 599-607, 2009.
- [79] HUANG, De-Shuang; YU, Hong-Jie. Normalized feature vectors: a novel alignment-free sequence comparison method based on the numbers of adjacent amino acids. **IEEE/ACM Transactions on Computational Biology and Bioinformatics**, v. 10, n. 2, p. 457-467, 2013.
- [80] CAI, Jie et al. Feature selection in machine learning: A new perspective. **Neurocomputing**, v. 300, p. 70-79, 2018.
- [81] WANG, Zhen; MOULT, John. SNPs, protein structure, and disease. **Human mutation**, v. 17, n. 4, p. 263-270, 2001.
- [82] LEACH, Andrew R.; LEACH, Andrew R. **Molecular modelling: principles and applications**. Pearson education, 2001.
- [83] KARPLUS, Martin; MCCAMMON, J. Andrew. Molecular dynamics simulations of biomolecules. **Nature structural biology**, v. 9, n. 9, p. 646-652, 2002.
- [84] RODRIGUEZ-BUSSEY, Isela G.; DOSHI, Urmi; HAMELBERG, Donald. Enhanced molecular dynamics sampling of drug target conformations. **Biopolymers**, v. 105, n. 1, p. 35-42, 2016.
- [85] LAZIM, Raudah; SUH, Donghyuk; CHOI, Sun. Advances in molecular dynamics simulations and enhanced sampling methods for the study of protein systems. **International journal of molecular sciences**, v. 21, n. 17, p. 6339, 2020.
- [86] BERNASCONI, Paul et al. A naturally occurring point mutation confers broad range tolerance to herbicides that target acetolactate synthase. **Journal of Biological Chemistry**, v. 270, n. 29, p. 17381-17385, 1995.
- [87] GUTTIERI, Mary J.; EBERLEIN, Charlotte V.; THILL, Donald C. Diverse mutations in the acetolactate synthase gene confer chlorsulfuron resistance in kochia (*Kochia scoparia*) biotypes. **Weed Science**, v. 43, n. 2, p. 175-178, 1995.
- [88] LI, Yan et al. In silico identification of genetic mutations conferring resistance to acetohydroxyacid synthase inhibitors: a case study of *Kochia scoparia*. **PloS one**, v. 14, n. 5, p. e0216116, 2019.
- [89] WHALEY, Cory M.; WILSON, Henry P.; WESTWOOD, James H. A new mutation in plant ALS confers resistance to five classes of ALS-inhibiting herbicides. **Weed Science**, v. 55, n. 2, p. 83-90, 2007.
- [90] KOLKMAN, Judith M. et al. Acetohydroxyacid synthase mutations conferring resistance to imidazolinone or sulfonylurea herbicides in sunflower. **Theoretical and Applied Genetics**, v. 109, n. 6, p. 1147-1159, 2004.
- [91] MASSA, D.; KRENZ, B.; GERHARDS, R. Target-site resistance to ALS-inhibiting herbicides in *Apera spica-venti* populations is conferred by documented and previously unknown mutations. **Weed Research**, v. 51, n. 3, p. 294-303, 2011.

[92] FOES, Matthew J. et al. A biotype of common waterhemp (*Amaranthus rudis*) resistant to triazine and ALS herbicides. **Weed Science**, v. 46, n. 5, p. 514-520, 1998.

[93] PATZOLDT, William L.; TRANEL, Patrick J. Multiple ALS mutations confer herbicide resistance in waterhemp (*Amaranthus tuberculatus*). **Weed science**, v. 55, n. 5, p. 421-428, 2007.

[94] BRANDT, Artur ML et al. Exploring the unbinding of *L. eishmania* (L.) amazonensis CPB derived-epitopes from H 2 MHC class I proteins. **Proteins: Structure, Function, and Bioinformatics**, v. 84, n. 4, p. 473-487, 2016.

[95] MCCOURT, J. A.; DUGGLEBY, R. G. Acetohydroxyacid synthase and its role in the biosynthetic pathway for branched-chain amino acids. **Amino acids**, v. 31, n. 2, p. 173-210, 2006.

[96] SCHLOSS, John V. Acetolactate synthase, mechanism of action and its herbicide binding site. **Pesticide Science**, v. 29, n. 3, p. 283-292, 1990.

[97] SHANER, Dale L.; SINGH, Bijay K.; STIDHAM, Mark A. Interaction of imidazolinones with plant acetohydroxyacid synthase: evidence for in vivo binding and competition with sulfometuron methyl. **Journal of agricultural and food chemistry**, v. 38, n. 5, p. 1279-1282, 1990.