

UNIVERSIDADE ESTADUAL PAULISTA “JULIO DE MESQUITA FILHO”

Programa de Pós-graduação em Ciência e Tecnologia de Materiais

Câmpus de Bauru

Augusto Cisconi Deienno

**ESTUDOS DE RELAÇÕES ESTRUTURA-REATIVIDADE-
ATIVIDADE DE COMPOSTOS ANTIVIRAIS α -CETOAMIDAS
VIA CÁLCULOS DE ESTRUTURA ELETRÔNICA**

BAURU

2023

Augusto Cisconi Deienno

**ESTUDOS DE RELAÇÕES ESTRUTURA-REATIVIDADE- ATIVIDADE
DE COMPOSTOS ANTIVIRAIS α -CETOAMIDAS VIA CÁLCULOS DE
ESTRUTURA ELETRÔNICA**

Dissertação apresentada como requisito à
obtenção do título de Mestre à Universidade
Estadual Paulista “Júlio de Mesquita Filho” -
Programa de Pós-Graduação em Ciência e
Tecnologia de Materiais, sob a orientação do Prof.
Dr. Augusto Batagin Neto

Bauru
2023

Deianno, Augusto Deianno.

ESTUDOS DE RELAÇÕES ESTRUTURA-REATIVIDADE-
ATIVIDADE DE COMPOSTOS ANTIVIRAIS α -CETOAMIDAS VIA
CÁLCULOS DE ESTRUTURA ELETRÔNICA/ Augusto Cisconi
Deianno, 2023

121 f. : il.

Orientador: Augusto Batagin Neto

Dissertação (Mestrado)-Universidade Estadual
Paulista (Unesp). Faculdade de Ciências, Bauru, 2023

1. Antivirais. 2. Índices de Reatividade. 3.
Relações Estrutura-Atividade I. Universidade Estadual
Paulista. Faculdade de Ciências. II. Título.

ATA DA DEFESA PÚBLICA DA DISSERTAÇÃO DE MESTRADO DE AUGUSTO CISCONI DEIENNO, DISCENTE DO PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA E TECNOLOGIA DE MATERIAIS, DA FACULDADE DE CIÊNCIAS - CÂMPUS DE BAURU.

Aos 28 dias do mês de abril do ano de 2023, às 14:00 horas, por meio de Videoconferência, realizou-se a defesa de DISSERTAÇÃO DE MESTRADO de AUGUSTO CISCONI DEIENNO, intitulada **Estudos de relações estrutura-reatividade-atividade de compostos antivirais α -cetoamidas via cálculos de estrutura eletrônica**. A Comissão Examinadora foi constituída pelos seguintes membros: Prof. Dr. AUGUSTO BATAGIN NETO (Orientador(a) - Participação Virtual) do(a) Coordenadoria de Curso de Engenharia de Produção / Instituto de Ciências e Engenharia - Unesp/ Campus de Itapeva, Prof. Dr. RAFAEL PLANA SIMÕES (Participação Virtual) do(a) Departamento de Bioprocessos e Biotecnologia / Faculdade de Ciências Agrônômicas - UNESP / Câmpus de Botucatu, Prof. Dr. PAULO NORONHA LISBOA FILHO (Participação Virtual) do(a) Departamento de Física / Faculdade de Ciências - Unesp/Câmpus de Bauru. Após a exposição pelo mestrando e arguição pelos membros da Comissão Examinadora que participaram do ato, de forma presencial e/ou virtual, o discente recebeu o conceito final APROVADO. Nada mais havendo, foi lavrada a presente ata, que após lida e aprovada, foi assinada pelo(a) Presidente(a) da Comissão Examinadora.

Prof. Dr. AUGUSTO BATAGIN NETO

Dedico este trabalho à minha mãe,
pelo apoio incondicional em toda a
minha trajetória acadêmica.

AGRADECIMENTOS

Agradeço ao professor Dr. Augusto Batagin Neto, meu orientador, por todo apoio na pesquisa.

Agradeço a todos os professores do Programa de Pós-Graduação em Ciência e Tecnologia de Materiais (POSMAT) e a Letícia Lopes Veronez, secretária do Programa, por toda a atenção e ajuda.

Agradeço a Capes-PROEX pela concessão da bolsa (88887.642942/2021-00) de mestrado.

Agradeço à minha família por todo apoio nos meus estudos.

Ao Grupo de Pesquisa "A inclusão da pessoa com deficiência, TGD/TEA ou superdotação e os contextos de aprendizagem e desenvolvimento" coordenado pela Professora Associada Vera Lúcia Messias Fialho Capellini, por toda a ajuda e apoio nessa jornada.

Ao Grupo do meu orientador, por toda a ajuda e apoio.

DEIENNO, A. C. **Estudos de relações estrutura-reatividade-atividade de compostos antivirais α -cetoamidas via cálculos de estrutura eletrônica.** 2023.

Dissertação (Mestrado em Ciência e Tecnologia de Materiais) - Universidade Estadual Paulista “Julio de Mesquita Filho” - UNESP, Bauru/SP, 2023.

RESUMO

Nos últimos anos a comunidade científica tem trabalhado intensamente na busca e desenvolvimento de novos fármacos que permitam inibir infecções virais (antirretrovirais), como a COVID-19. De fato, um grande número de compostos tem sido avaliado, em especial, por meio de técnicas de busca sistematizada via triagem virtual (*virtual screening*), contudo, as metodologias envolvidas apresentam limitações intrínsecas, principalmente associadas à determinação apropriada de propriedades eletrônicas, estruturais e de reatividade local de forma concomitante. Neste estudo, técnicas de modelagem molecular baseadas na teoria do funcional da densidade foram empregadas na investigação destas propriedades em compostos antivirais da família das α -cetoamidas no sentido de se melhor compreender mecanismos de ação e guiar trabalhos experimentais futuros. Vinte e sete derivados de α -cetoamidas, cuja atividade antiviral contra coronavírus tem sido reportada, foram avaliados. Os resultados permitem identificar características de interesse para a obtenção de compostos ativos, em especial nota-se que sistemas mais promissores devem apresentar baixa tendência de interação com agentes eletrofílicos sobre as regiões centrais da cadeia, ao mesmo tempo em que apresentam alta reatividade sobre regiões terminais específicas. Análises preliminares de descritores 3D baseados no potencial eletrostático reforçam os resultados obtidos. Com base nestes resultados uma série de novos derivados cetoamídicos foram propostos e avaliados, resultando em compostos com alta atividade antiviral prevista. De forma geral, os dados apresentados permitem avaliar a potencialidade de diferentes compostos por meio de cálculos de estrutura eletrônica de baixo custo computacional.

Palavras-chave: Antivirais. Índices de Reatividade. Relações Estrutura-Atividade.

ABSTRACT

In recent years the scientific community has worked intensively in the search and development of new drugs to suppress viral infections (antiretroviral drugs), such as COVID-19. In fact, a large number of compounds have been tested, especially through systematized search techniques via virtual screening, but the involved methodologies have intrinsic limitations, mainly associated with the appropriate description of electronic, structural, and local reactivity properties concomitantly. In this study, molecular modeling techniques weaknesses based on density functional theory were employed to investigate these properties in antiviral compounds of the α -ketoamide family in order to better understand mechanisms of action and guide future experimental work. Twenty-seven derivatives of α -ketoamides, whose antiviral activity against coronaviruses has been reported, were evaluated. The results allow the identification of features of interest for obtaining active compounds, in particular it is noted that the most promising systems should show low tendency to interact with electrophilic agents over the central regions of the chain, while showing high reactivity over specific terminal regions. Preliminary analyzes of 3D descriptors based on electrostatic potentials reinforce the obtained results. Based on these results a series of new ketoamide derivatives have been proposed and evaluated, resulting in compounds with predicted high antiviral activity. Overall, the data presented allow the evaluation of the potentiality of different compounds by means of low computational cost electronic structure calculations.

Keywords: Antivirals. Reactivity Indexes. Structure-Activity Relationships.

LISTA DE ILUSTRAÇÕES

Figura 1. Estrutura típica do coronavírus.....	22
Figura 2. Representação da reprodução do coronavírus.....	23
Figura 3. Variantes do SARS-CoV 2.....	25
Figura 4. Procedimento de desenvolvimento das Vacinas contra o SARS-CoV 2.	28
Figura 5. Design de inibidores de α -cetoamidas.....	31
Figura 6. Estudo de docagem molecular demonstrando a interação M_{pro} -Inibidor de dois derivados de α -cetoamidas (11a1 e 11b1) no sítio de interação.....	31
Figura 7. Esquema reacional para a síntese de derivados de α -cetoamidas.....	32
Figura 8. Ilustração da obtenção de derivados de α -cetoamidas.....	32
Figura 9. Mecanismo proposto de ação de inibitória covalente das α -cetoamidas atuando sobre resíduos específicos da enzima do vírus SARS-CoV 2. b) Exemplificação da atividade inibitória para os compostos 11a e GC373.....	34
Figura 10. Workflow dos métodos de Triagem Virtual, TBVS e LBVS.....	37
Figura 11. Estrutura básica dos compostos avaliados.....	41
Figura 12. a) Estrutura básica e regiões dos compostos avaliados. b) alterações realizadas sobre os as regiões P1, P2 e P3 para a obtenção dos grupos de testagem A, B e D. c) alterações realizadas sobre os as regiões P1, P2 e P3 para a obtenção dos grupos de testagem B, C e D.....	46
Figura 13. Estrutura bidimensional do Paxlovid para testagem do grupo E.....	50
Figura 14. Telas de configuração do software Gabedit empregadas no estudo conformacional via DM.....	52
Figura 15. Ilustração do processo de construção de superfícies reativas em compostos com estruturas diferentes.....	59
Figura 16. Distribuições de energia dos confôrmeros dos compostos a) 11a1, b)	

11b1, c) 11a2, d) 11b2, e) 11c, f) 11d, g) 11e, h) 11g obtidas do estudo estrutural62

Figura 17. Distribuições de energia dos confôrmeros dos compostos a) 11h, b) 11i, c) 11j, d) 11k, e) 11m, f) 11n, g) 11o, h) 11p obtidas do estudo estrutural.....63

Figura 18. Distribuições de energia dos confôrmeros dos compostos a) 11q, b) 11r, c) 11s, d) 11t, e) 11u, f) 13a, g) 13b, h) 14a obtidas do estudo estrutural.....64

Figura 19. Distribuições de energia dos confôrmeros dos compostos a) I2, b) N1, c) N3, obtidas do estudo estrutural.....65

Figura 20. Estruturas de menor energia obtidas do estudo estrutural após otimização via abordagem DFT/B3LYP/6-31G(d), sendo: a) 11a1 b) 11a2 c) 11b1 d) 11b2 e) 11c f) 11d g) 11e h) 11g i) 11h j) 11i k) 11j l) 11k.....68

Figura 21. Estruturas de menor energia obtidas do estudo estrutural após otimização via abordagem DFT/B3LYP/6-31G(d), sendo a) 11m b) 11n c) 11o d) 11p e) 11q f) 11r g) 11s h) 11t i) 11u j) 13a k) 13b l) 14b.....69

Figura 22. Estruturas de menor energia obtidas do estudo estrutural após otimização via abordagem DFT/B3LYP/6-31G(d), sendo a) I2 b) N1 c) N3.....70

Figura 23. IFCA, PEM e Centróides dos inibidores: 11a1, 11a2, 11b1, 11b2, 11c, 11d, 11e, 11g, 11h e 11i.....71

Figura 24. IFCA e PEM dos inibidores: 11j, 11k, 11m, 11n, 11o, 11p, 11q, 11r, 11s e 11t.....72

Figura 25. IFCA, PEM e Centróides dos inibidores: 11u, 13a, 13b, 14b, I2, N1 e N373

Figura 26. Influência da distância, d , entre centróides de f^* e f na classificação de compostos com atividade antiviral (círculos vermelhos definem compostos ativos $\ln(IC50) < 0.0$, enquanto que círculos em preto indicam compostos de menor atividade antiviral).....75

Figura 27. Predição da atividade biológica em escala logarítmica nos eixos x e y76

Figura 28. Predição da atividade biológica em escala logarítmica dos novos

sistemas moleculares investigados.....80

Figura 29. Influência da distância, d , entre centróides de f^* e f na classificação de compostos com atividade antiviral dos grupos de testagem (círculos vermelhos definem compostos ativos $\ln(IC50) < 0.0$, enquanto que círculos em preto indicam compostos de menor atividade antiviral).....82

Figura 30. Representação das Superfícies Moleculares 3D com os descritores IFCA e PEM: 11a1 e 11a2. Os eixos coordenados x , y e z , representam as coordenadas espaciais em Å (omitidas por simplicidade).....84

Figura 31. Representação das Superfícies Moleculares 3D com os descritores IFCA e PEM: 11b1 e 11b2. Os eixos coordenados x , y e z , representam as coordenadas espaciais em Å (omitidas por simplicidade).....85

Figura 32. Representação das Superfícies Moleculares 3D com os descritores IFCA e PEM: 11m e 11n. Os eixos coordenados x , y e z , representam as coordenadas espaciais em Å (omitidas por simplicidade).....86

Figura 33. Representação das Superfícies Moleculares 3D com a identificação de sítios relevantes identificados via análises de *Random Forest*. O composto 11b1 é ilustrado para fins de identificação das regiões P1, P2 e P3.....88

Figura A1. Mutações do SARS-CoV 2: D614G, V483a, S943P.....96

Figura A2. Mutações do SARS-CoV 2: E484K, N501Y, L452.....96

Figura A3. Mutações do SARS-CoV 2: Q677, E484Q, S477G/N.....96

Figura A4. Representação das Superfícies Moleculares 3D com os descritores IFCA e PEM: 11c e 11d. Os eixos coordenados x , y e z , representam as coordenadas espaciais em Å (omitidas por simplicidade).....100

Figura A5. Representação das Superfícies Moleculares 3D com os descritores IFCA e PEM: 11e e 11g. Os eixos coordenados x , y e z , representam as coordenadas espaciais em Å (omitidas por simplicidade).....101

Figura A6. Representação das Superfícies Moleculares 3D com os descritores IFCA e PEM: 11h e 11i. Os eixos coordenados x , y e z , representam as coordenadas espaciais em Å (omitidas por simplicidade).....102

Figura A7. Representação das Superfícies Moleculares 3D com os descritores IFCA e PEM: 11j e 11k. Os eixos coordenados x, y e z, representam as coordenadas espaciais em Å (omitidas por simplicidade).....	103
Figura A8. Representação das Superfícies Moleculares 3D com os descritores IFCA e PEM: 11o e 11p. Os eixos coordenados x, y e z, representam as coordenadas espaciais em Å (omitidas por simplicidade).....	104
Figura A9. Representação das Superfícies Moleculares 3D com os descritores IFCA e PEM: 11q e 11r. Os eixos coordenados x, y e z, representam as coordenadas espaciais em Å (omitidas por simplicidade).....	105
Figura A10. Representação das Superfícies Moleculares 3D com os descritores IFCA e PEM: 11s e 11t. Os eixos coordenados x, y e z, representam as coordenadas espaciais em Å (omitidas por simplicidade).....	106
Figura A11. Representação das Superfícies Moleculares 3D com os descritores IFCA e PEM: 11u e 13a. Os eixos coordenados x, y e z, representam as coordenadas espaciais em Å (omitidas por simplicidade).....	107
Figura A12. Representação das Superfícies Moleculares 3D com os descritores IFCA e PEM: 13b e 14b. Os eixos coordenados x, y e z, representam as coordenadas espaciais em Å (omitidas por simplicidade).....	108
Figura A13. Representação das Superfícies Moleculares 3D com os descritores IFCA e PEM: I2 e N1. Os eixos coordenados x, y e z, representam as coordenadas espaciais em Å (omitidas por simplicidade).....	109
Figura A14. Representação da Superfície Molecular 3D com os descritores IFCA e PEM: N3. Os eixos coordenados x, y e z, representam as coordenadas espaciais em Å (omitidas por simplicidade).....	110
Figura A15. IFCA, PEM e Centróides dos compostos do grupo de testagem: A1-A7.....	111
Figura A16. IFCA, PEM e Centróides dos compostos do grupo de testagem: B1-B8.....	112
Figura A17. IFCA, PEM e Centróides dos compostos do grupo de testagem: C1-	

C9.....	113
Figura A18. IFCA, PEM e Centróides dos compostos do grupo de testagem: D1-D7.....	114
Figura A19. IFCA, PEM e Centróides dos compostos do grupo de testagem: D8-D14.....	115
Figura A20. IFCA, PEM e Centróides do Paxlovid do grupo de testagem E.....	115
Figura A21. Representação bidimensional da posição dos centróides associados à f^+ (verde) e f^- (magenta) em relação dos compostos do grupo de testagem: A1-A7.....	116
Figura A22. Representação bidimensional da posição dos centróides associados à f^+ (verde) e f^- (magenta) em relação dos compostos do grupo de testagem: B1-B8.....	117
Figura A23. Representação bidimensional da posição dos centróides associados à f^+ (verde) e f^- (magenta) em relação dos compostos do grupo de testagem: C1-C9.....	118
Figura A24. Representação bidimensional da posição dos centróides associados à f^+ (verde) e f^- (magenta) em relação dos compostos do grupo de testagem: D1-D7.....	119
Figura A25. Representação bidimensional da posição dos centróides associados à f^+ (verde) e f^- (magenta) em relação dos compostos do grupo de testagem: D8-D14.....	120
Figura A26. Representação bidimensional da posição dos centróides associados à f^+ (verde) e f^- (magenta) em relação ao Paxlovid do grupo E de testagem.....	120

LISTA DE TABELAS

Tabela 1. Plataformas vacinais.....	27
Tabela 2. Identificação dos compostos avaliados, valores de IC ₅₀ (μM) e referências experimentais: 11a1-11g.....	42
Tabela 3. Identificação dos compostos avaliados, valores de IC ₅₀ e referências experimentais: 11h-11p.....	43
Tabela 4. Identificação dos compostos avaliados, valores de IC ₅₀ (μM) e referências experimentais: 11q-14b.....	44
Tabela 5. Identificação dos compostos avaliados, valores de IC ₅₀ e referências experimentais: I2, N1 e N3.....	45
Tabela 6. Compostos propostos para testagem dos grupos A e C.....	47
Tabela 7. Compostos propostos para testagem do grupo B.....	48
Tabela 8. Compostos propostos para testagem do grupo D.....	49
Tabela 9. Descritores eletrônicos considerados nas análises multivariadas.....	55
Tabela 10. Faixa de energia da diferença entre o confôrmero mais e menos estável	66
Tabela 11. Confôrmeros selecionados de menor energia.....	67
Tabela 12. Atividade antiviral prevista para os compostos do grupo de testagem “A”	78
Tabela 13. Atividade antiviral prevista para os compostos do grupo de testagem “B”	78
Tabela 14. Atividade antiviral prevista para os compostos do grupo de testagem “C”	79
Tabela 15. Atividade antiviral prevista para os compostos do grupo de testagem “D”	79
Tabela 16. Atividade antiviral prevista para o Paxlovid, desenvolvido pela Pfizer..	79

Tabela A1. Representação bidimensional da posição dos centróides associados à f^+ (verde) e f^- (magenta) em relação a estrutura dos compostos: 11a1, 11a2, 11b1, 11c, 11d, 11e, 11g, 11h, 11i.....97

Tabela A2. Representação bidimensional da posição dos centróides associados à f^+ (verde) e f^- (magenta) em relação a estrutura dos compostos: 11j, 11k, 11m, 11n, 11o, 11p, 11q, 11r, 11s, 11t.....98

Tabela A3. Representação bidimensional da posição dos centróides associados à f^+ (verde) e f^- (magenta) em relação a estrutura dos compostos: 11u, 13a, 13b, 14b, 12, N1, N3.....99

LISTA DE ABREVIATURAS E SIGLAS

3CL _{pro}	Protease do SARS-Cov 2
ACE2	Enzima conversora de angiotensina-2
AE	Afinidade eletrônica
AMBER	Família de campos de força obtida via refinamento de energia, do inglês <i>assisted model building with energy refinement</i>
A _S	Área molecular superficial total
B3LYP	Becke, 3 parâmetros, Lee-Yang-Parr (funcional)
Ch _k	Carga líquida de Hirshfeld associada ao átomo <i>k</i> do <i>core</i> (com <i>k</i> = 1 a 19)
Chelp	Método de partição de cargas a partir do potencial eletrostático molecular, do inglês <i>charges from electrostatic potential</i>
C _q	Componente q do centróide associado aos IFCA (q = x, y ou z)
COVID-19	Doença associada ao coronavírus definida em 2019, do inglês <i>coronavirus disease 2019</i>
d ^{CX-CG}	Distância entre os centróides associados a f ^x e o centro geométrico da molécula (X= + ou -)
DFT	Teoria do funcional da densidade, do inglês <i>density functional theory</i>
DM	Dinâmica molecular
D _q	Componente q do momento de dipolo elétrico (q = x, y, ou z)
D _T	Módulo do momento de dipolo total
E	Envelope do SARS-Cov 2
E _{HOMO}	Energia do último orbital ocupado (via orbitais de Kohn-Sham)

E_{LUMO}	Energia do primeiro orbital desocupado (via orbitais de Kohn-Sham)
E_{T}	Energia Total
f_k^X	IFCA associado ao átomo k do <i>core</i> (com $k = 1$ a 19 e $X = +$ ou $-$)
GAFF	Campo de força AMBER geral, do inglês <i>general AMBER force field</i>
HF	Hartree-Fock (método)
HIV	Vírus da Imunodeficiência Humana
IC_{50}	Concentração necessária para inibir 50% da atividade enzimática, do inglês <i>50% inhibitory concentration</i>
IFCA	Índices de Fukui Condensados aos Átomos
M	Membrana do SARS-Cov 2
MERS-CoV	Coronavírus responsável pela síndrome respiratória do oriente médio, do inglês <i>Middle East respiratory syndrome coronavirus</i>
MMFF94	Campo de força Merck
M_{pro}	Outra denominação da protease 3CL_{pro}
mRNA	RNA mensageiro
MSE	Erro quadrático médio, do inglês <i>mean squared error</i>
N	Nucleocapsídeo
OMS	Organização Mundial da Saúde
OOB	Amostras não selecionadas, do inglês <i>out-of-bag</i>
ψ_i	Potencial eletrostático molecular associado ao i -ésimo ponto da superfície discretizada
PC	Componente principal, do inglês <i>principal component</i>
PCA	Análise de componentes principais, do inglês <i>principal component</i>

analysis

PCM	Modelo de solvatação contínuo polarizável, do inglês <i>polarizable continuum model</i>
PEM	Potencial Eletrostático Molecular
PI	Potencial de ionização
PM6	Método paramétrico 6, do inglês <i>parametric method 6</i>
Pot f_i^X	Potenciais de Fukui ($X = +, -$ e 0) associados ao i -ésimo ponto da superfície molecular discretizada
QSAR	Relação quantitativa entre estrutura e atividade, do inglês <i>quantitative structure-activity relationship</i>
R_A	Índice de aceitação de elétrons
R_D	Índice de doação eletrônica
RF	Floresta aleatória, do inglês <i>random forest</i>
RLM	Regressão linear múltipla
RLS	Regressão linear simples
RM-RF	Ranking médio associado às análises via RF
RNA	Ácido ribonucleico, do inglês <i>ribonucleic acid</i>
S	Espiga SARS-Cov 2, do inglês <i>Spike</i>
SARS-CoV	Coronavírus associado à síndrome respiratória aguda grave, do inglês <i>severe acute respiratory syndrome - coronavirus</i>
SARS-CoV 2	Coronavírus associado à síndrome respiratória aguda grave, do inglês <i>severe acute respiratory syndrome - coronavirus 2</i>
VBM	Variante a ser monitorada, do inglês <i>Variant Being Monitored</i>
VOI	Variante de Interesse, do inglês <i>Variant of Interest</i>

VOC	Variante de Preocupação, do inglês <i>Variant of Concern</i>
VOHC	Variante de Alta Consequência, do inglês <i>Variant of High Consequence</i>
V_T	Volume molecular total
ω^-	Capacidade eletrodoadora
ω^+	Capacidade eletroaceitora
WESP	World Economic Situation and Prospects

SUMÁRIO

1. INTRODUÇÃO.....	19
1.1. COVID-19.....	19
1.1.1. Pandemia e seus impactos.....	19
1.1.2. O Coronavírus.....	21
1.1.3. Vacinas e tratamento.....	25
1.2. Inibidores e seus mecanismos de ação.....	29
1.3. Design de fármacos antivirais.....	35
1.4. Motivação do estudo.....	38
2. OBJETIVOS.....	40
2.1. Objetivos gerais.....	40
2.2. Objetivos específicos.....	40
3. MATERIAIS E MÉTODOS.....	41
3.1. Materiais.....	41
3.1.1. Inibidores α -cetoamidas avaliados na literatura.....	41
3.1.2. Inibidores α -cetoamidas propostos.....	45
3.2. Métodos.....	50
3.2.1. Estudos conformacionais	50
3.2.2. Estudo de reatividade.....	53
3.2.3. Obtenção de descritores eletrônicos.....	54
3.2.4. Análises estatísticas e estabelecimento de correlações e modelos preditivos..	57
4. RESULTADOS.....	62
4.1. Estudo estrutural.....	62
4.2. Estudo de reatividade.....	70
4.3. Estudos de QSAR.....	75
4.4. Proposição e prospecção de novos compostos.....	77
4.5. Análise 3D de descritores de reatividade.....	83
5. CONCLUSÕES.....	89
REFERÊNCIAS.....	91
ANEXO A.....	96
APÊNDICE I.....	97
APÊNDICE II.....	100
APÊNDICE III.....	111
APÊNDICE IV.....	116

1. INTRODUÇÃO

1.1. COVID-19

Este capítulo versa sobre a pandemia e seus impactos, morfologia do coronavírus, as vacinas e tratamentos.

1.1.1. Pandemia e seus impactos

Em dezembro de 2019 na cidade de Wuhan, província de Hubei, na China, um surto de pneumonia causada por um novo vírus, denominado SARS-CoV 2 (do inglês *severe acute respiratory syndrome coronavirus 2*), responsável por uma síndrome respiratória aguda grave, se espalhou para muitos países de forma acelerada, tornando-se um problema de saúde global, classificado pela Organização Mundial de Saúde (OMS) como uma pandemia (CHAN et al., 2020; ZHU et al., 2020).

O sequenciamento genômico demonstrou que este patógeno (SARS-CoV 2) possui 79,6% da sequência genética de outro já conhecido, o SARS-CoV (DAI et al., 2020; ZHOU et al., 2020). A infecção deste vírus leva ao desenvolvimento da doença denominada COVID-19 (do inglês *coronavirus disease 2019*), a qual, em humanos se apresenta de três formas: *i*) casos assintomáticos; *ii*) indivíduos que apresentam sintomas leves, semelhantes a de um resfriado comum (febre, tosse, secreção respiratória, dores de cabeça, entre outros), e *iii*) indivíduos que apresentam a forma grave da doença, casos em que os pulmões ficam comprometidos e outros sintomas mais graves podem surgir como disfunção

cardíaca, hepática e renal, os quais podem eventualmente evoluir para a sepse (XAVIER et al., 2020).

A transmissão do vírus se dá pelo contato com gotículas de saliva expelidas no ar. Ainda não existem medicamentos antivirais específicos para este novo coronavírus.

Em especial, a pandemia do COVID-19 desencadeou a nível global uma interrupção de oferta e demanda de produtos e serviços, causando um grande choque na economia mundial. A redução de trabalho, o fechamento de empresas, a baixa produtividade e o distanciamento social imposto visando o controle do contágio levaram ao bloqueio de fornecimento, ao mesmo tempo em que a perda de renda e as demissões em massa, consequências da quarentena, reduziram o consumo e investimento (ACOSTA, 2021).

Conforme relata o World Economic Situation and Prospects (WESP) (2021, p.1) com relação aos impactos econômicos mundiais causados pela doença em 2020:

“Em 2020, a produção mundial diminuiu 4,3%, três vezes mais do que durante a crise financeira global de 2009. A modesta recuperação de 4,7%, prevista para 2021, dificilmente compensaria as perdas sofridas em 2020. A pandemia atingiu mais duramente as economias desenvolvidas, com um declínio de produção estimado em 5,6% em 2020, devido às rigorosas e prolongadas medidas de bloqueio que foram impostas em muitos países europeus e em algumas partes dos Estados Unidos durante o surto.” (tradução livre).

De fato, muita incerteza em relação ao coronavírus ainda permanece, principalmente com o surgimento de novas variantes resistentes às vacinas atualmente desenvolvidas.

Segundo a OMS, conforme estudo do Instituto Butantan: “[...] o número de mortes por COVID-19 no mundo pode ser o triplo do que os registros oficiais indicam, chegando a 18,2 milhões de pessoas até o final de 2021, contra os 5,9

milhões divulgados oficialmente” (INSTITUTO BUTANTAN, 2022). Estima-se que o número de mortos seja três vezes maior do que o número oficial.

Segundo Wang et al. (2022), diversas razões interferiram para que ocorresse esse grande número de subnotificações, é preciso o teste com resultado positivo para COVID-19 para que se efetive o registro do óbito; os sistemas de registros são diferentes em cada região ou país; há uma discordância médica mundial em declarar a morte por COVID-19 quando a pessoa estava infectada pelo SARS-CoV 2 e também os fatores políticos que impediram as devidas notificações em diversas regiões. É fato a ocorrência de muitas mortes consequentes de várias outras doenças durante as medidas de isolamento impostas, que aumentaram o impacto do número de mortos durante a pandemia.

1.1.2. O Coronavírus

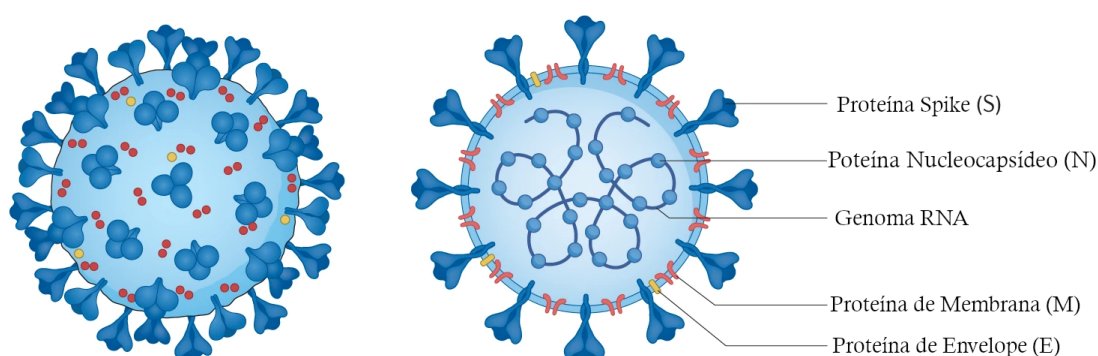
Coronaviridae é uma família de vírus divididos em *alphacoronavirus*; *beta* coronavírus; *gama* coronavírus e *delta* coronavírus. São agentes capazes de infectar aves e mamíferos. O coronavírus que infecta os humanos pertence à classe de beta coronavírus e é geralmente associado a doenças respiratórias como resfriados comuns (HARTENIAN; 2020).

Nos últimos anos, ao menos três coronavírus capazes de infectar animais selvagens foram transmitidos para o ser humano, causando doenças respiratórias mais graves do que um resfriado comum. O SARS-CoV foi identificado em 2002, sendo também responsável por uma síndrome respiratória aguda grave. Em 2011 o MERS-CoV (coronavírus responsável pela síndrome respiratória do oriente médio, do inglês *Middle East respiratory syndrome coronavirus*) foi identificado na Arábia

Saudita espalhando-se para vários outros países. Em 2019, foi identificado o SARS-CoV 2 (HARTENIAN; 2020).

O genoma do coronavírus consiste de uma fita de RNA positivo, o qual é envolvido pela proteína de nucleocapsídeo (N) e uma membrana composta pelas proteínas: spike S, membrana M e envelope E. A proteína S é necessária para a entrada viral, ligando-se à célula-alvo, sendo responsável por iniciar o processo de fusão com a membrana da célula hospedeira. A proteína M viabiliza a incorporação de componentes virais essenciais em novos vírions (partículas virais completas) durante a morfogênese. Por fim, a proteína E define um canal iônico na membrana do vírus, participando também da montagem viral (HARTENIAN; 2020). A Figura 1, ilustra a estrutura típica do coronavírus.

Figura 1. Estrutura típica do coronavírus

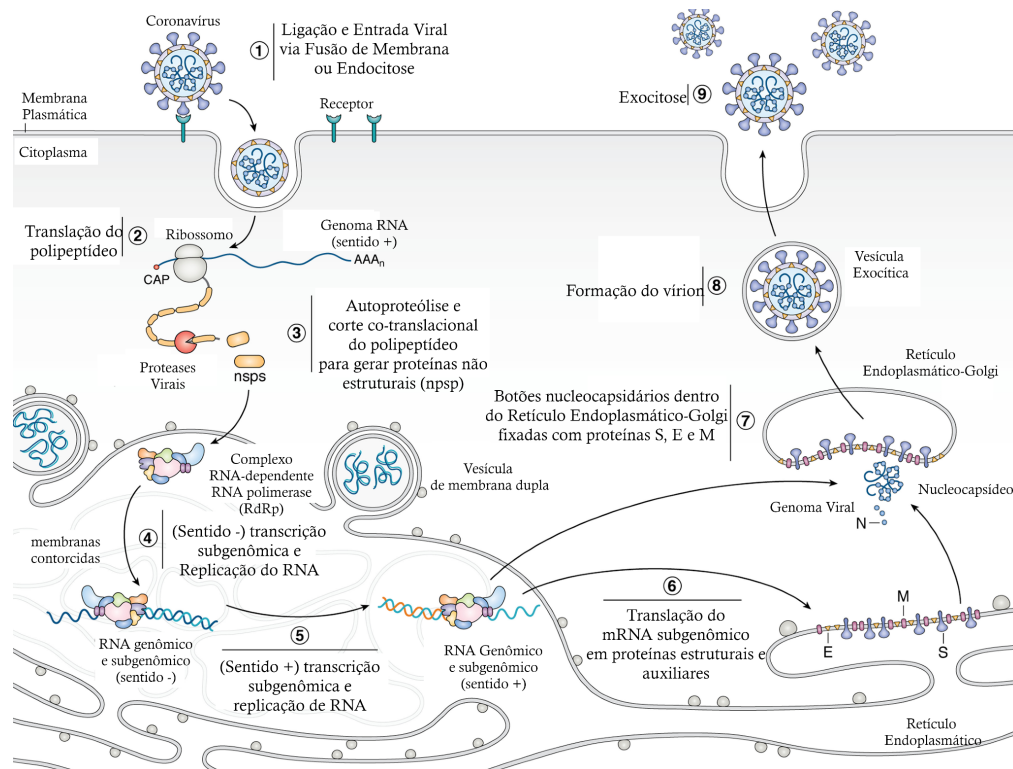


Fonte: Adaptado de Lamers e Haagmans (2022)

Em particular, a proteína S se acopla a enzima conversora de angiotensina-2 (ACE2), e, por meio de divisões, promove a fusão da membrana viral com a célula hospedeira (processo de endocitose). Essa proteína possui dois domínios, S1 e S2, sendo por meio de S1 que o receptor se liga ao hospedeiro. S2, por sua vez, promove a fusão da membrana permitindo o acesso do vírus ao citoplasma da

célula. A Figura 2 representa o mecanismo de entrada, replicação e saída do coronavírus na célula hospedeira.

Figura 2. Representação da reprodução do coronavírus



Fonte: Adaptado de Hartenian (2020)

Nota-se que o processo de entrada e saída do coronavírus na célula hospedeira se dá em 9 etapas: *i)* ligação e entrada viral via fusão de membrana ou endocitose; *ii)* translação de polipeptídeo; *iii)* auto proteólise e corte co-translacional de polipeptídeo para gerar proteínas não estruturais; *iv)* sentido negativo de transcrição subgenômica e replicação do RNA; *v)* sentido positivo de transcrição subgenômica e replicação do RNA; *vi)* translação de RNA mensageiro (mRNA) subgenômico em proteínas estruturais e proteínas auxiliares; *vii)* botões nucleocapsidários em um compartimento intermediário de retículo

endoplasmático-Golgi fixado com proteínas estruturais S, E e M; *viii*) formação do vírus; e *ix*) exocitose (HARTENIAN, 2020).

Em especial, o SARS-CoV 2 possui uma protease (3CL_{pro}), também chamada de M_{pro}, que tem um papel essencial na replicação viral. Ela atua na quebra de proteínas virais, promovendo a replicação de RNA, diferentemente da proteína S, que se liga na célula hospedeira. Devido a isso, a 3CL_{pro} define um importante alvo de atuação de inibidores, impedindo a replicação e proliferação do vírus, e desta forma reduzindo a carga viral e a gravidade da doença a ele associada (CHUCK et al., 2011).

A taxa de mutação do coronavírus, em geral, é muito elevada, ocorrendo na fase da replicação. Deste modo, quanto maior o número de pessoas infectadas, maior a possibilidade de surgirem novas variantes do vírus. Em especial, no caso do SARS-CoV 2, nota-se uma taxa de mutação relativamente menor em relação a outros vírus, principalmente devido à ação de enzimas restauradoras que atuam em possíveis falhas do processo de replicação. Em geral, um estado de atenção é requerido quando elevados picos de contaminação são observados, em especial, envolvendo o comprometimento de vias respiratórias, o que, acaba por facilitar a transmissibilidade do patógeno (CANTÓN, 2021).

A Figura 3 ilustra as variantes do SARS-CoV 2, suas linhagens, a classificação de cada variante e sua data de classificação. A classificação vem do inglês “*Variant Being Monitored (VBM)*”, “*Variant of Interest (VOI)*”, “*Variant of Concern (VOC)*” e “*Variant of High Consequence (VOHC)*”.

Figura 3. Variantes do SARS-CoV 2

Designação da OMS	Linagem	VOC	Data de designação VOI	VBM
Alfa	B.1.1.7 e linhagem Q	29 de dezembro de 2020		21 de setembro de 2021
Beta	B.1.351 e linhagens descendentes	29 de dezembro de 2020		21 de setembro de 2021
Gama	P.1 e linhagens decedentes	29 de dezembro de 2020		21 de setembro de 2021
Delta	B.1.617.2 e linhagem AY	15 de junho de 2020		21 de setembro de 2021
Epsilon	B.1.427, B.1.429	19 de março de 2020	26 de fevereiro de 2021, 29 de junho de 2021	21 de setembro de 2021
Eta	B.1.525		26 de fevereiro de 2021	21 de setembro de 2021
Iota	B.1.526		26 de fevereiro de 2021	21 de setembro de 2021
Kappa	B.1.617.1		7 de maio de 2021	21 de setembro de 2021
N/A	B.1.617.3		7 de maio de 2021	21 de setembro de 2021
Zeta	P.2		26 de fevereiro de 2021	21 de setembro de 2021
Mu	B.1.621, B.1.621.1			21 de setembro de 2021

Fonte: Adaptado de *Centers for Disease control and Prevention* (2022)

Tais variantes definem-se principalmente devido à presença de mutações na proteína S. Detalhes acerca destas mutações são brevemente descritos por Cosar et al. (2021) e sumarizadas no Anexo A.

No Brasil, em especial, foi relatado o surgimento da variante P.1, a qual teve grande impacto no número de óbitos, principalmente na região norte do país no período entre 4 dezembro de 2020 e 2 de janeiro de 2021 (NAVECA, 2021).

1.1.3. Vacinas e tratamento

Os métodos que são empregados para a produção de vacinas, são bastante diversificados. Três métodos são utilizados atualmente para produção: vacinas de subunidades inativadas e proteicas; vacinas que utilizam vetores virais e vacinas de mRNA.

As vacinas de subunidades inativadas e proteicas são produzidas a partir da cultura inativada de vírus, inativadas quimicamente (CREECH, et.al. 2021). De

acordo com a OMS, são 29 vacinas em ensaios clínicos que utilizam a tecnologia de subunidade proteica (fragmentos de proteínas do coronavírus). O mecanismo de ação dessas vacinas concentra-se na proteína S, parte do vírus que penetra na célula hospedeira. (FERRAZ, 2021) Comumente utiliza-se a química na melhora da resposta imune adicionando substâncias que estimulam o sistema imunológico na fórmula das vacinas, essas substâncias são adjuvantes químicos que melhoram a resposta imune, segundo Diniz e Ferreira (2010, apud FERRAZ, 2021).

As vacinas que utilizam vetores virais empregam vírus modificados incapazes de se reproduzir (em geral adenovírus). Esses vírus geneticamente modificados carregam a sequência gênica do vírus de interesse. Essa técnica pode prejudicar a imunização se o indivíduo já tiver alguma imunidade ao adenovírus. No caso da COVID-19, elas impedem que a doença se apresente em sua forma mais grave, não impedindo a contaminação e aparecimento de sintomas leves. Em geral oferecem proteção a novas variantes (CREECH, et.al., 2021).

As vacinas mRNA utilizam o mRNA para transportar o material genético do patógeno. Nessa técnica avançada, nanopartículas lipídicas são usadas como agentes protetores da proteína S estabilizada, codificando o mRNA para o espaço intracelular. Assim, o hospedeiro utiliza o mRNA para produzir a proteína alvo (proteína S), induzindo uma resposta imune coordenada. Essa vacina possui alta eficácia e poucas reações adversas. Ela pode ser produzida rapidamente e produz respostas em curto intervalo de tempo (CREECH, et.al. 2021).

A Tabela 1 apresenta as plataformas vacinais divulgadas no ano de 2020.

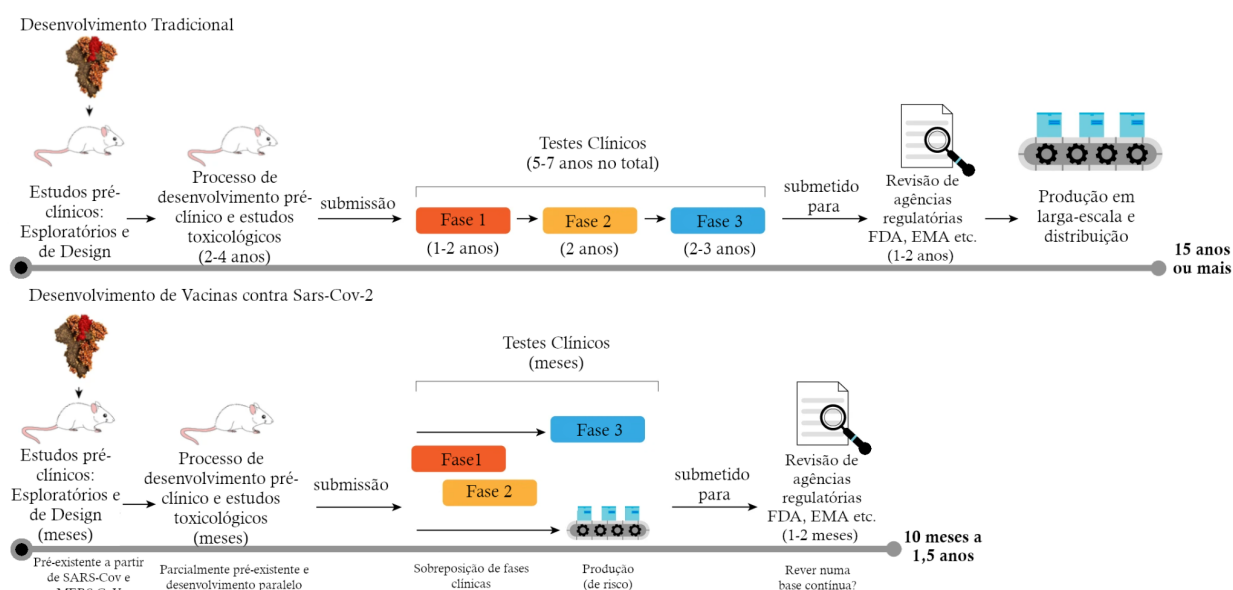
Tabela 1. Plataformas vacinais

Vetor viral	Genéticas-RNA	Proteicas	Virais inativadas
AstraZeneca / Oxford (Reino Unido) Adenovírus de Chipanzé (ChAd)	Pfizer / BioNTech (Alemanha) RNA mensageiro	Novavax (Estados Unidos) Sub-unidades	Sinovac (China)
CanSino Biological (China) Adenovírus humano tipo 5 (Ad5)	Moderna (Estados Unidos) RNA mensageiro		Sinopharm Wuhan (China)
Gamaleya (Rússia) Adenovírus humano tipos 5 e 26 (Ad26 e Ad5)			Sinopharm Beijin (China)
Jansen (Estados Unidos) Adenovírus humano tipo 26 (Ad26)			

Fonte: Lima, Almeida e Kfoury (2021)

No que tange o desenvolvimento de vacinas para o enfrentamento da COVID-19, vale ressaltar o intenso esforço e curto prazo para a obtenção de produtos com alta capacidade de proteção. Para fins de ilustração, a Figura 4 apresenta um diagrama de desenvolvimento de vacinas tradicionalmente observado e o impulsionado pela pandemia.

Apesar da preocupação global em acelerar o desenvolvimento de vacinas no período da pandemia, há ainda a necessidade de se encontrar fármacos antivirais para o combate e tratamento desta doença e outras similares, o que ainda é um foco de pesquisa de diversos grupos.

Figura 4. Procedimento de desenvolvimento das Vacinas contra o SARS-CoV 2

Fonte: Adaptado de Krammer (2020)

Em especial, pesquisadores e estudiosos em saúde pública mundial compartilham uma grande preocupação associada ao surgimento de novas pandemias (BALMANT e FONTANETTO, 2022). Neste contexto, para antever possíveis ameaças, é crucial o monitoramento genômico e epidemiológico de agentes infecciosos. De fato, investimentos em sistemas de saúde fortalecidos e produção de vacinas tornam um país preparado para enfrentar emergências sanitárias, similares às enfrentadas nos anos de 2020-2022. A destruição ambiental e o aumento populacional tendem a intensificar o risco de transmissão de doenças infecciosas entre animais e seres humanos, sendo desta forma essencial garantir um amplo conjunto de medidas e normas sanitárias para prevenir e controlar doenças já conhecidas e aquelas que possam surgir (BALMANT e FONTANETTO, 2022).

De fato, muitos especialistas concordam que a humanidade se encontra em constante risco de sofrer uma nova pandemia com o surgimento de um vírus

ignorado pela ciência ou até mesmo desconhecido. As últimas pandemias que a humanidade já sofreu, como a gripe espanhola em 1918; a gripe asiática em 1957; a SARS-COV em 2003 até a SARS-COV-2 em 2019, evidenciam a grande vulnerabilidade dos sistemas de vigilância e combate a novos agentes infecciosos (INSTITUTO BUTANTAN, 2021).

Alguns especialistas, por exemplo, sugerem que a COVID-19 pode não ser a pior pandemia já enfrentada pela humanidade. Sendo previstas pandemias futuras similares que poderão levar a um número ainda maior de mortes e impactos secundários significativos. Tais projeções levam em conta a ocorrência de mutações em vírus que tipicamente se transmitem por meio da via respiratória, como o H5N1 (considerado, inclusive, um potencial causador de futuras grandes pandemias), principalmente por sua forma de transmissibilidade (MANIR, 2020). Tal cenário evidencia a necessidade de estudos contínuos acerca do desenvolvimento de vacinas, fármacos e inibidores para o tratamento destas doenças (INSTITUTO BUTANTAN, 2021).

1.2. Inibidores e seus mecanismos de ação

Neste capítulo, aborda os fármacos testados e a discussão a respeito sobre trabalhos publicados sobre os inibidores investigados, bem como os mecanismos de ação reportados.

Dentre eles, o *Remdesivir*, medicamento antiviral inicialmente desenvolvido para tratamento do Ebola (WANG et al., 2020) foi aprovado para uso emergencial em 1 de maio de 2020 (modificado em 28 de agosto) nos EUA, sendo liberado para o tratamento em pessoas hospitalizadas. (BEIGEL et al., 2020)

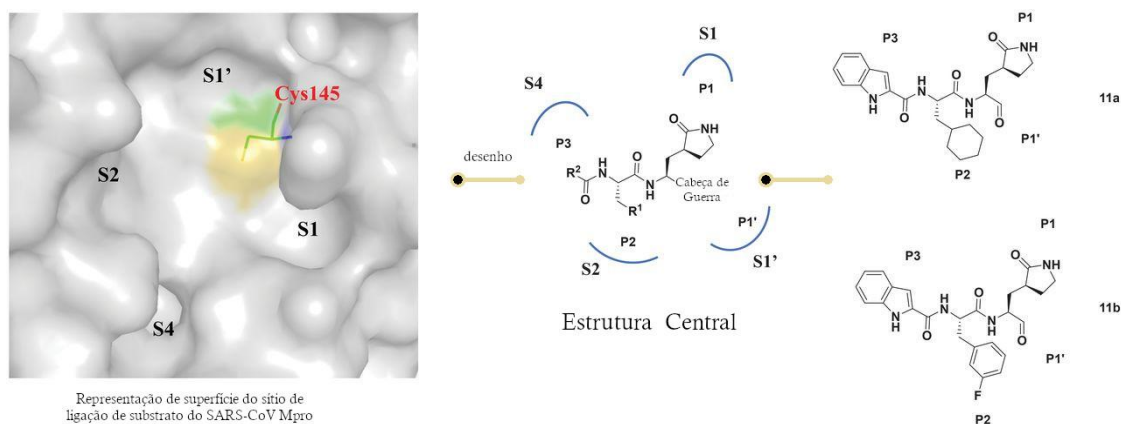
Tal processo foi repetido em vários países. Testes clínicos, no entanto, sugeriram que o uso deste medicamento de forma isolada não seria suficiente para todos os pacientes, apresentando resultados mais promissores quando associado a outros medicamentos com ação de resposta imune como o *baricitinib* (BEIGEL et al., 2020). Combinações baseadas em *a-interferon* e *lopinavir-ritonavir* (anti-HIV) também foram testadas, apresentando, no entanto, resultados pouco promissores e fortes efeitos colaterais, além de alta toxicidade (CAO et al., 2020; DAI et al., 2020).

Em outra linha de investigação, o uso de fármacos amplamente testados no tratamento de outras doenças foi também avaliado para o tratamento da COVID-19. Em especial, o uso da cloroquina e hidroxicloroquina, comumente empregados no tratamento da Malária e do Lúpus, causou uma grande polêmica em muitos países, inclusive no Brasil (FIOROTTO DE MATTOS, 2021). Tais compostos, derivados da quinina, foram testados *in vitro* contra o SARS-CoV 2, apresentando um resultado eficaz contra a infecção, porém não demonstraram a eficácia esperada em testes *in vivo* (GROSSE et al., 2021). De forma geral, em estudos realizados em pacientes com COVID-19, não foi evidenciado uma melhora significativa dos indivíduos tratados com estes medicamentos em relação aos que não receberam tratamento. Por não apresentarem eficácia evidenciada no combate à COVID-19, a OMS declarou não haver segurança no uso destes medicamentos no combate à doença, contudo, esses medicamentos continuaram a ser usados em diversos países (FIOROTTO DE MATTOS, 2021).

Conforme apontado por Dai et. al. (2020), trabalhos que visam o desenvolvimento de inibidores contra o coronavírus estão bastante focados na inibição da proteína M_{pro} , presente na proteína S do SARS-CoV 2. Em especial estes autores têm proposto o uso de compostos derivados de α -cetoamidas como

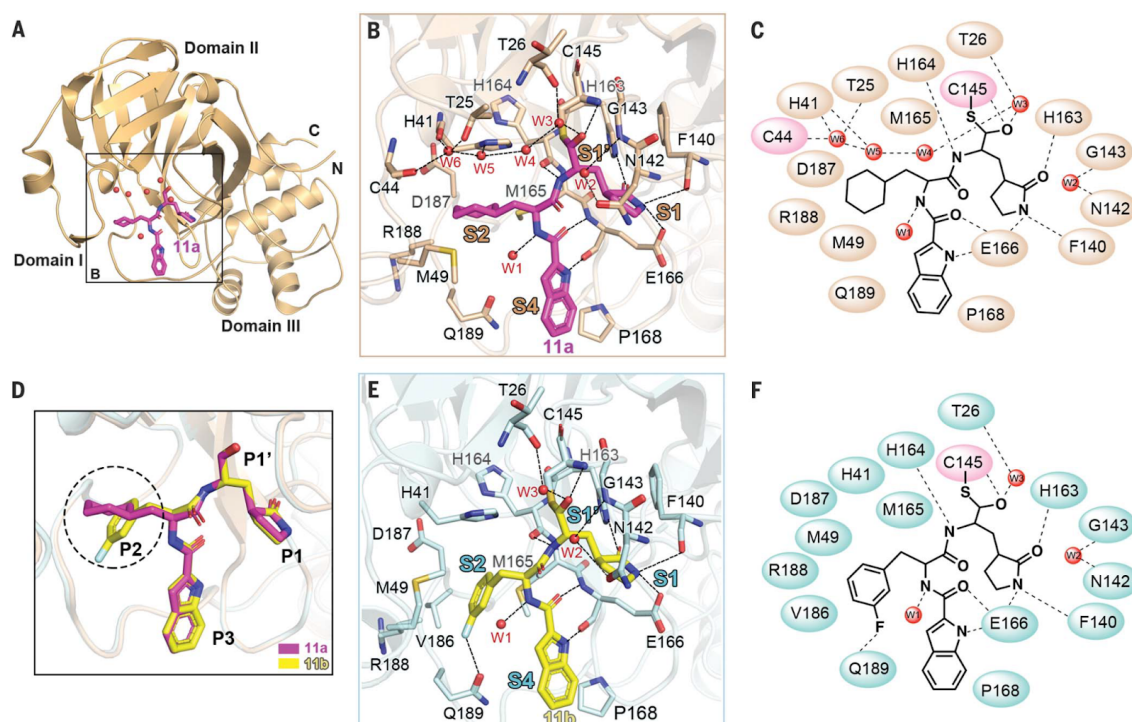
potenciais inibidores, os quais possuem uma grande flexibilidade de síntese e alta interação com sítios ativos destes alvos, conforme ilustram as Figuras 5 e 6.

Figura 5. Design de inibidores de α -cetoamidas



Fonte: Dai et al. (2020)

Figura 6. Estudo de docagem molecular demonstrando a interação M_{pro} -Inibidor de dois derivados de α -cetoamidas (11a1 e 11b1) no sítio de interação.



Fonte: Dai et al. (2020)

A Figura 7 demonstra o esquema reacional comumente empregado na síntese de derivados de α -cetoamidas (ZHANG, 2020a), enquanto a Figura 8 ilustra a composição de derivados típicos com distintos ligantes laterais (11r, 13a, 13b e 14b).

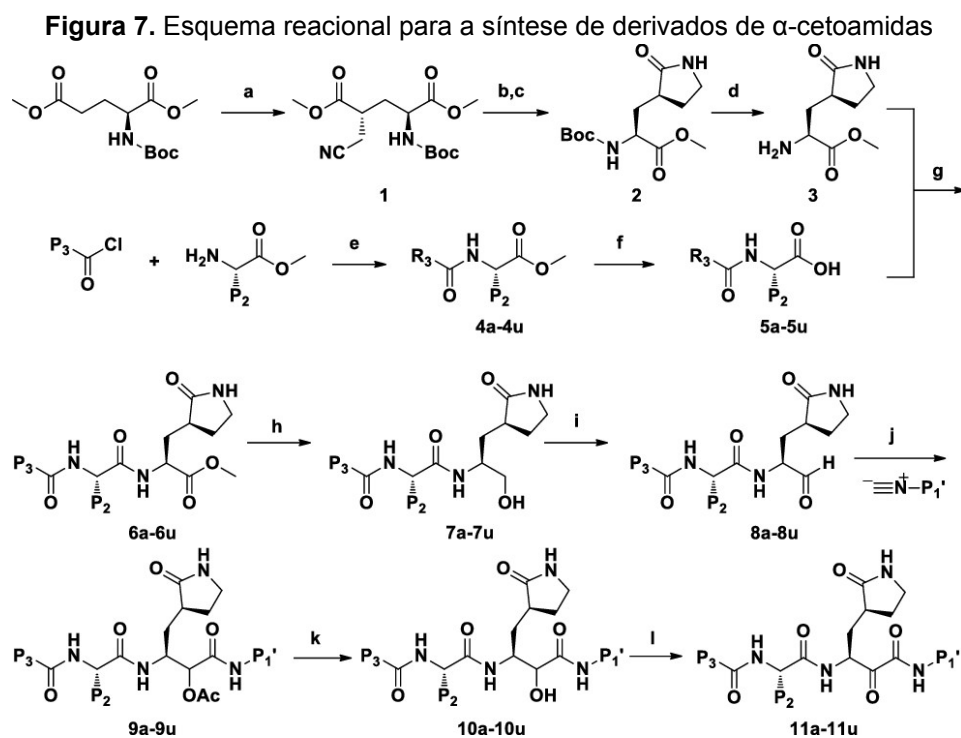
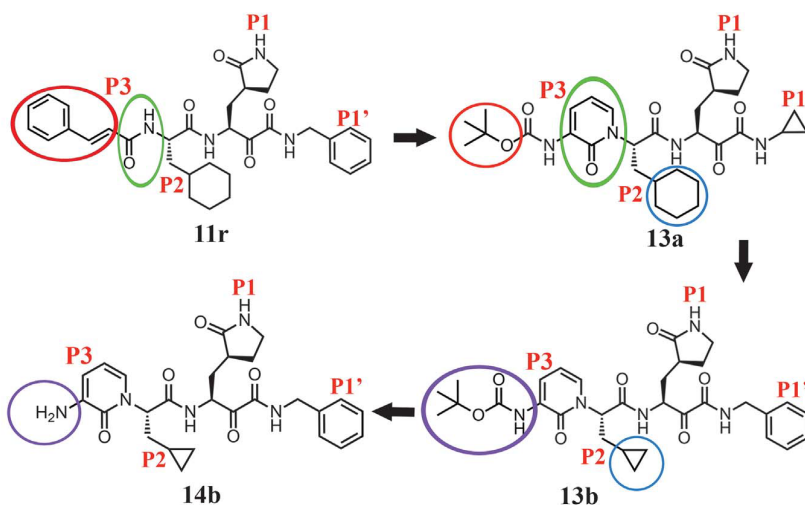


Figura 8. Ilustração da obtenção de derivados de α -cetoamidas



Nota-se que o esquema de reação das α -cetoamidas permite identificar regiões de modificação estrutural para obtenção de um inibidor mais promissor. Em especial os compostos 11a e 11b, segundo DAI et al. (2020), apresentaram propriedades antivirais promissoras, demonstrando alta capacidade de inibição da M_{pro} do SARS-CoV 2, sendo o composto 11a (Figura 7a) selecionado para futuros estudos clínicos.

Zhang et al. (2020a), reportaram alta potência inibitória do composto 11r (Figura 9) contra o SARS-CoV. Segundo os autores, tal composto também tem potencial contra SARS-CoV 2. Dentre a gama de compostos estudados por eles, em especial as estruturas 11r, 11s e 11n tiveram melhor resposta contra o SARS-CoV. Estudos conduzidos por Zhang et al. (2020b) buscaram sintetizar novas estruturas, contra o SARS-CoV 2 a partir do composto promissor 11r, sintetizando 13a, 13b e 14b. Respostas promissoras no que tange a farmacocinética foram obtidas para o composto 13b, sugerindo a possibilidade de administração direta nos pulmões.

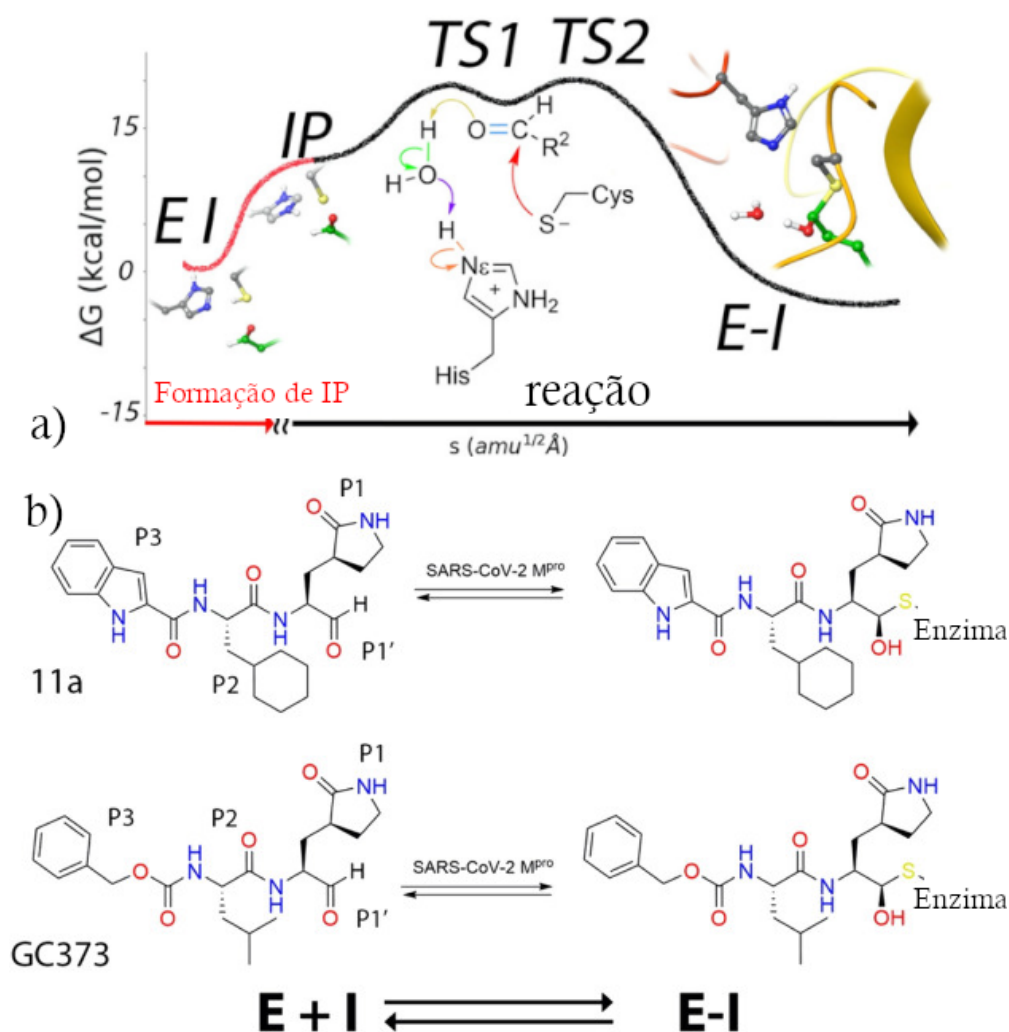
Tais dados evidenciam a potencialidade destes compostos no enfrentamento à doença, indicando um conjunto promissor de estruturas químicas a serem investigadas.

Possíveis mecanismos de ação de inibidores baseados em α -cetoamidas sobre a enzima do vírus tem sido descritos por Ramos-Guzmán; Ruiz-Pernía; Tuñón (2021, p.1):

“A reação (entre o fármaco e a enzima) é iniciada com a formação do par de íons díades catalíticos após uma transferência de prótons entre os resíduos cisteína (Cys145) para histidina (His41). Deste estado ativado, a inibição covalente prossegue com o ataque nucleofílico do átomo desprotonado (enxofre da Cys145) sobre o átomo de carbono do aldeído do fármaco, bem como uma transferência protônica mediada por água do átomo de nitrogênio da His41 para o átomo de oxigênio do aldeído (fármaco)”;

a Figura 9 ilustra esse processo:

Figura 9. a) Mecanismo proposto de ação covalente dos inibidores (I) α -cetoamidas atuando sobre resíduos específicos da enzima (E) do vírus SARS-CoV 2 via a formação de pares iônicos (IP) e estados de transição (TS1 e TS2). b) Exemplificação da atividade inibitória para os compostos 11a e GC373.



Fonte: Adaptado de Ramos-Guzmán; Ruiz-Pernía; Tuñón (2021)

Tais dados sugerem a relevância de processos de transferência de carga entre os inibidores e os alvos virais os quais ainda carecem de maiores estudos.

1.3. *Design* de fármacos antivirais

Dentre as diferentes técnicas de *design* de compostos antivirais, destacam-se os métodos de triagem virtual (ou *virtual screening*). De modo geral, o objetivo desta técnica situa-se na identificação de compostos bioativos mais promissores a partir de uma vasta biblioteca de estruturas químicas. Em geral tais métodos baseiam-se em 2 abordagens distintas: a) triagem virtual baseada na estrutura do alvo molecular, a qual envolve estudos de interação de possíveis compostos bioativos com alvos moleculares proteicos de relevância para o desenvolvimento viral; e b) triagem virtual baseada em características do ligante a qual tem como foco principal a realização de estudos comparativos envolvendo estruturas químicas conhecidamente bioativas com outros compostos de potencial interesse.

A abordagem envolvendo a estrutura do alvo molecular é baseada em estudos de *docking*. Nesses estudos os ligantes são posicionados no sítio ativo do alvo molecular de interesse biológico, sendo realizada uma busca posicional e conformacional. Complexo ligante-alvo mais prováveis são identificados com base na afinidade de posições (orientações) relativas entre o ligante e o sítio-receptor, estimadas de acordo com parâmetros de interação eletrostáticas, químicos e/ou semi empíricos, que definem funções de rankeamento denominadas *score*. Com base nesses parâmetros, compostos bioativos podem ser identificados dentro de uma vasta biblioteca de dados guiando trabalhos experimentais subsequentes. Apesar da potencialidade da técnica de triagem virtual baseada na estrutura do alvo molecular, a mesma apresenta uma forte dependência com os métodos e algoritmos empregados na predição dos modos de ligação e na estimativa da afinidade entre ligante e alvo molecular. Em especial, características associadas à

flexibilidade/rigidez dos compostos ligantes, bem como conformação dos sítios ativos proteicos definem parâmetros relevantes nos resultados deste tipo de triagem.

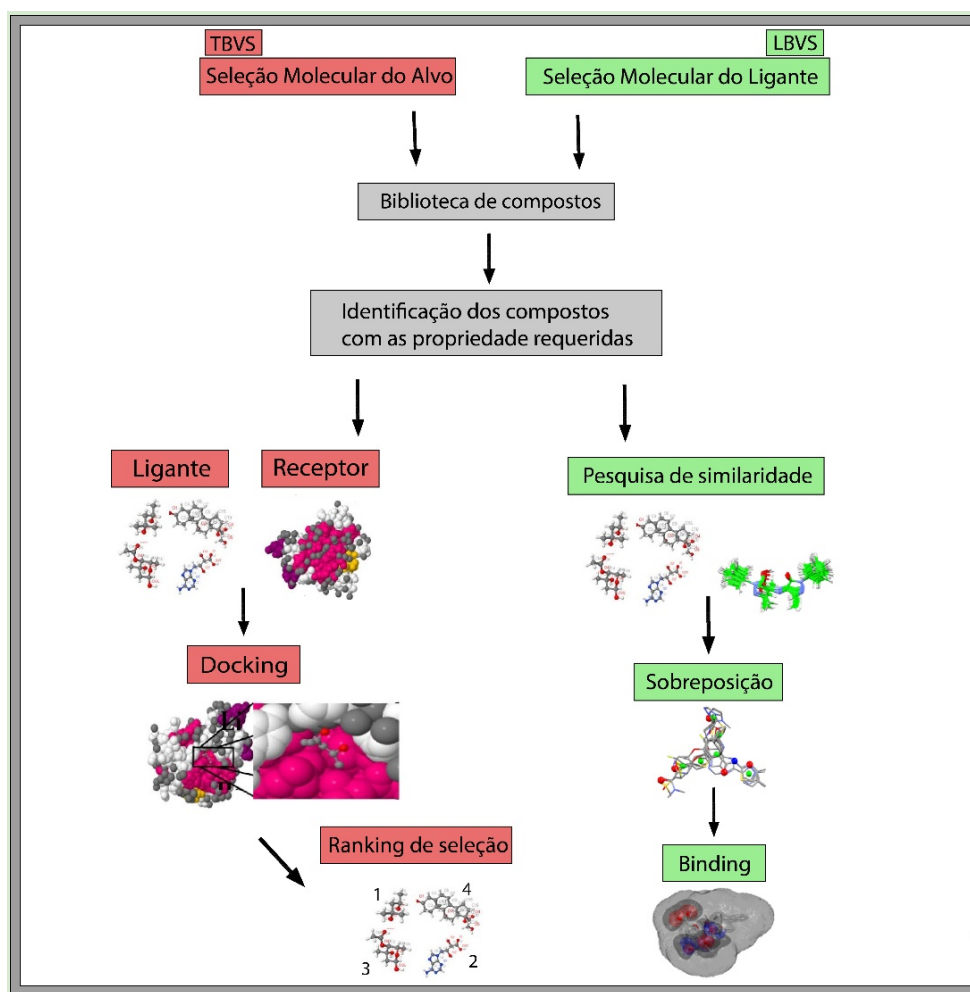
A abordagem envolvendo a triagem virtual baseada em características do ligante, diferentemente da técnica anterior, não leva em consideração a estrutura tridimensional do receptor biológico. Ao invés disso, características estruturais e propriedades diversas de compostos de interesse (oriunda tipicamente de bancos de dados) são estimadas e comparadas com as advindas de compostos conhecidamente ativos. Dentre os métodos mais utilizados nesta abordagem destacam-se: a) busca baseada em padrão farmacofórico; b) busca por similaridade química e estrutural; c) estabelecimentos e relações quantitativas entre estrutura e atividade biológica (QSAR); e d) busca por meio de métodos de aprendizado de máquina. A primeira abordagem leva em conta parâmetros estruturais e análise comparativa entre a posição relativa de grupos funcionais específicos com o mesmo papel químico (eletrófilos, nucleófilos, grupos hidrofóbicos, grupos aromáticos etc.), conhecidos como farmacóforos, na busca por compostos promissores similares à fármacos com atividade já reportada. Na abordagem por similaridade, apenas características estruturais são consideradas, o grau de similaridade com compostos ativos é geralmente estimado com base em diferentes coeficientes de similaridade, sendo método muito dependente da escolha desses parâmetros. Tal abordagem pressupõe que moléculas estruturalmente similares tendem a apresentar propriedades físico-químicas e atividades biológicas semelhantes. Abordagens baseadas em QSAR utilizam um conjunto inicial de moléculas, que podem ser modeladas por softwares específicos, ou podem ser obtidas através de base de dados diversas, para realização de um estudo de análise conformacional visando a identificação de conformações viáveis (ou seja, de menor energia). Descritores

eletrônicos, estruturais e topológicos diversos podem ser utilizados para estudos estatísticos e obtenção de uma equação de predição de atividade biológica (geralmente por meio de regressões lineares).

Por fim, abordagens baseadas em técnicas de aprendizado de máquina, utilizam um grande conjunto de informações moleculares para treinamento e delineamento de modelos capazes de diferenciar compostos ativos e inativos.

Para fins de exemplificação, na Figura 10, são ilustradas características básicas representativas das diferentes abordagens empregadas na triagem virtual de fármacos.

Figura 10. Workflow dos métodos de Triagem Virtual, TBVS e LBVS.



Fonte: Silva et al. (2021)

1.4. Motivação do estudo

A busca por novos compostos ativos unicamente por tentativa e erro define-se atualmente como uma abordagem muito pouco efetiva, principalmente devido aos altos custos de produção esperados e a dificuldade de obtenção de compostos com atividade de fato otimizada.

Neste sentido, estudos teóricos e computacionais definem-se como alternativas indispensáveis na seleção de compostos promissores. Em especial simulação e modelagem computacional via técnicas de química quântica podem contribuir de forma significativa na identificação de compostos ativos e direcionamento de trabalhos experimentais, reduzindo o número de estudos sínteses necessárias (YANG et al., 2005). De fato, a obtenção de regras preditivas baseadas em descritores eletrônicos pode guiar a síntese e prospecção de compostos com características desejáveis, considerando-se propriedades químicas diversas dos compostos.

Grande parte dos estudos teóricos de modelagem e prospecção de fármacos baseia-se na identificação de sítios ativos, envolvendo a análises de *docking* molecular (com sítios alvo) ou a análise de propriedades eletrônicas intrínsecas dos compostos isolados. De forma geral, são raros os estudos envolvendo uma análise concomitante de propriedades eletrônicas, estruturais e de reatividade dos sistemas, sendo que dados estruturais acabam, por muitas vezes, dominando as análises realizadas.

No presente estudo propõe-se a investigação de dados de estrutura eletrônica de uma família de compostos derivados de α -cetoamidas com atividade biológica já evidenciada (DAI et al., 2020; ZHANG et al. 2020a; ZHANG et al. 2020b; YANG et al., 2005), atrelando-os a dados estruturais, no sentido de se identificar

processos de interação relevantes e propor compostos com propriedades antivirais otimizadas. Os resultados obtidos permitem identificar características desejáveis à derivados de com atividade antiviral intensificada o que pode guiar a síntese de novos compostos otimizados.

A presente dissertação se encontra dividida da seguinte forma: i) no Capítulo 1 apresenta-se a introdução contendo uma breve contextualização e levantamento da literatura; ii) no Capítulo 2 são apresentados os objetivos gerais e específicos do trabalho; iii) no Capítulo 3 são apresentados os materiais estudados e aspectos principais vinculados às metodologias de estudo empregadas; iv) no Capítulo 4 são apresentados os resultados obtidos do estudo estrutural, de reatividade, de desenvolvimento de ferramenta de avaliação 3D e de relações estrutura atividade para os diferentes compostos; v) no Capítulo 5 são apresentadas as principais conclusões do trabalho, seguido das referências. São apresentados nos apêndices resultados adicionais e nos anexos informações introdutórias complementares.

2. OBJETIVOS

2.1. Objetivos gerais

O presente trabalho tem como principal objetivo identificar características eletrônicas, estruturais e de reatividade relevantes na atividade antiviral de compostos do tipo α -cetoamidas, tidos como promissores no tratamento da COVID-19.

2.2. Objetivos específicos

São objetivos específicos do presente trabalho:

- Realização do estudo estrutural de conjuntos de moléculas;
- Cálculo e análise de diferenciados descritores eletrônicos;
- Cálculo e análise de índices de reatividade (índices de Fukui condensados aos átomos, análise da posição de centróides de reatividade, avaliação do grau de maciez química e sua similaridade com amino-ácidos) e potencial eletrostático molecular;
- Buscar padrões de correlação linear (por meio de análise de coeficiente de regressão linear de Pearson) entre os dados de estrutura eletrônica e atividade antiviral;
- Proposição e avaliação de novos sistemas promissores.

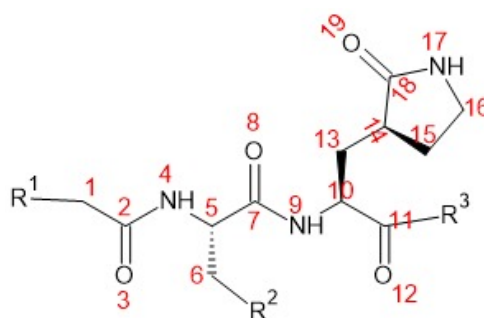
3. MATERIAIS E MÉTODOS

3.1. Materiais

3.1.1. Inibidores α -cetoamidas avaliados na literatura

A Figura 11 ilustra a estrutura básica comum a todos os compostos avaliados. Os grupos laterais R^1 - R^3 ilustram a posição de distintas ramificações conforme reportado por Dai et al (2020), Zhang et al. (2020a); Zhang et al. (2020b); e Yang et al., (2005). Com base nestes trabalhos, um total de 27 derivados foram selecionados, considerando-se a existência de uma estrutura básica comum (*core*).

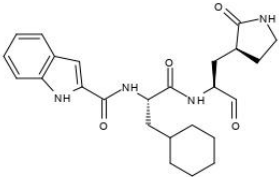
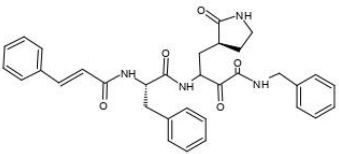
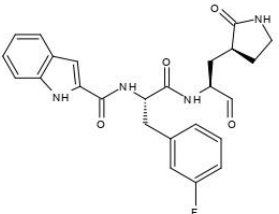
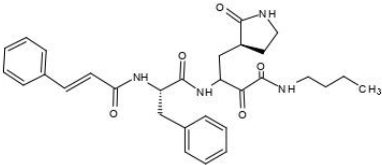
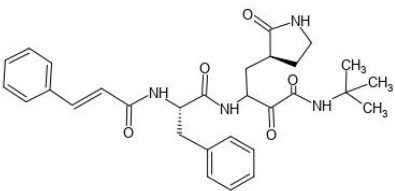
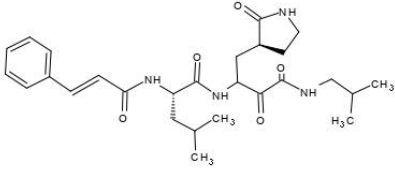
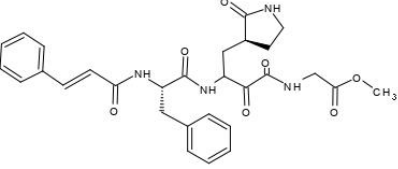
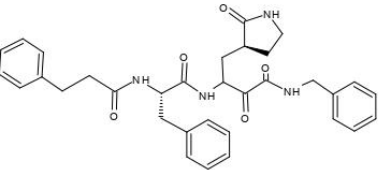
Figura 11. Estrutura básica dos compostos avaliados



Fonte: Autoria própria.

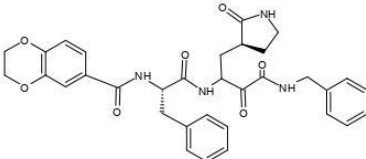
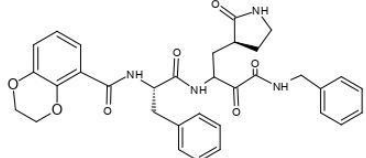
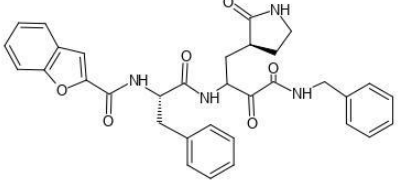
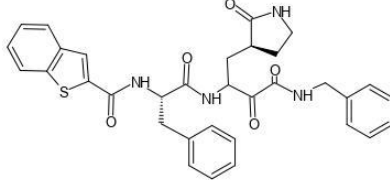
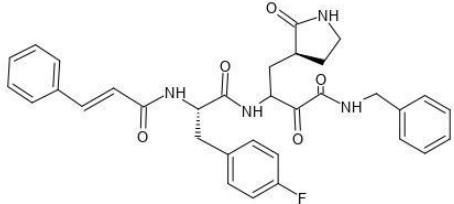
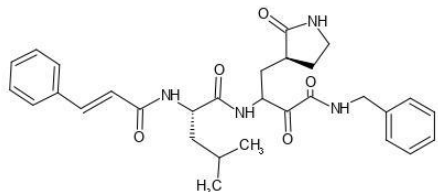
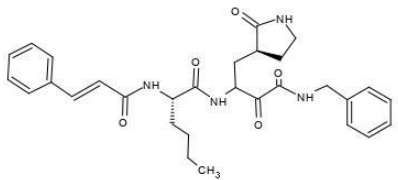
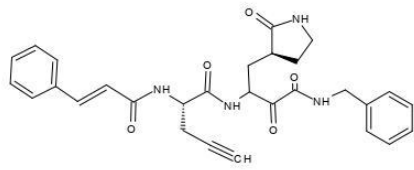
As Tabelas 2-5 ilustram as estruturas dos distintos compostos avaliados com os respectivos ligantes R^1 , R^2 e R^3 , bem como valores de IC_{50} (concentração em μM necessária do fármaco para inibir 50% da atividade enzimática nos testes *in vitro*) e as referências dos trabalhos experimentais. Os compostos 14b, I2, N1 e N3 não possuem valores de IC_{50} reportados.

Tabela 2. Identificação dos compostos avaliados, valores de IC₅₀ (μM) e referências experimentais:
11a1-11g.

composto	Estrutura	IC ₅₀	Referência
11a1		0.053	(DAI, W. et al., 2020.)
11a2		1.950	(ZHANG, L. et al., 2020a.)
11b1		0.040	(DAI, W. et al., 2020.)
11b2		50.000	(ZHANG, L. et al., 2020a.)
11c		43.190	(ZHANG, L. et al., 2020a.)
11d		11.370	(ZHANG, L. et al., 2020a.)
11e		4.470	(ZHANG, L. et al., 2020a.)
11g		3.590	(ZHANG, L. et al., 2020a.)

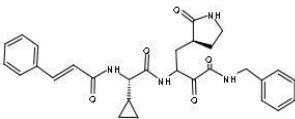
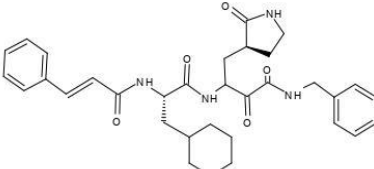
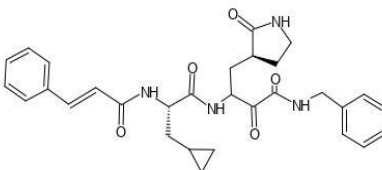
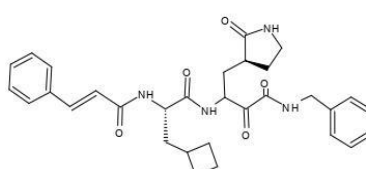
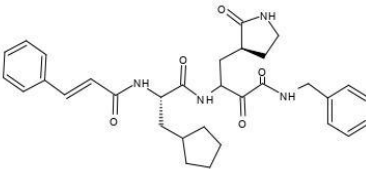
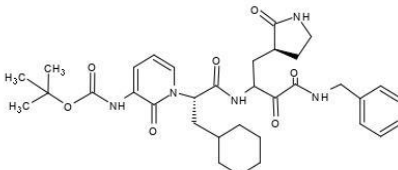
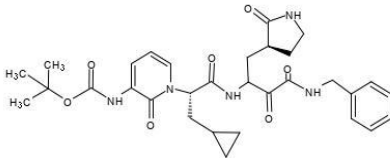
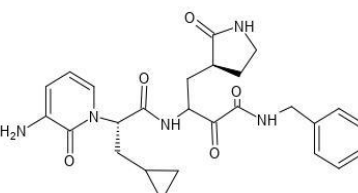
Fonte: Autoria própria

Tabela 3. Identificação dos compostos avaliados, valores de IC₅₀ e referências experimentais:
11h-11p.

composto	Estrutura	IC ₅₀	Referência
11h		10.730	(ZHANG, L. et al., 2020a.)
11i		11.880	(ZHANG, L. et al., 2020a.)
11j		16.710	(ZHANG, L. et al., 2020a.)
11k		9.680	(ZHANG, L. et al., 2020a.)
11m		50.000	(ZHANG, L. et al., 2020a.)
11n		0.330	(ZHANG, L. et al., 2020a.)
11o		8.500	(ZHANG, L. et al., 2020a.)
11p		10.680	(ZHANG, L. et al., 2020a.)

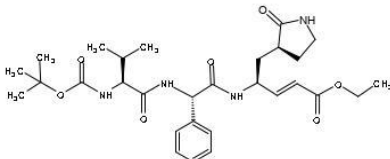
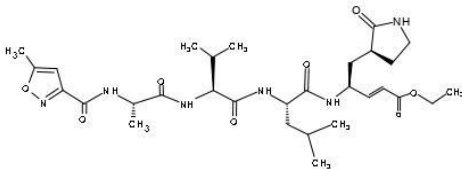
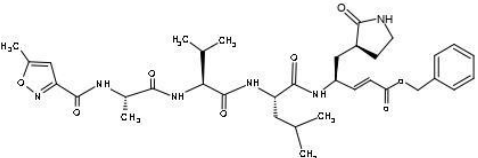
Fonte: Autoria própria

Tabela 4. Identificação dos compostos avaliados, valores de IC₅₀ (μM) e referências experimentais:
11q-14b.

composto	Estrutura	IC ₅₀	Referência
11q		6.270	(ZHANG, L. et al., 2020a.)
11r		0.710	(ZHANG, L. et al., 2020a.)
11s		0.240	(ZHANG, L. et al., 2020a.)
11t		1.440	(ZHANG, L. et al., 2020a.)
11u		1.270	(ZHANG, L. et al., 2020a.)
13a		2.390	(ZHANG, L. et al., 2020b.)
13b		0.670	(ZHANG, L. et al., 2020b.)
14b		---	(ZHANG, L. et al., 2020b.)

Fonte: Autoria própria

Tabela 5. Identificação dos compostos avaliados, valores de IC₅₀ e referências experimentais: I2, N1 e N3.

composto	Estrutura	IC ₅₀	Referência
I2		---	(YANG, H. et al., 2005.)
N1		---	(YANG, H. et al., 2005.)
N3		---	(YANG, H. et al., 2005.)

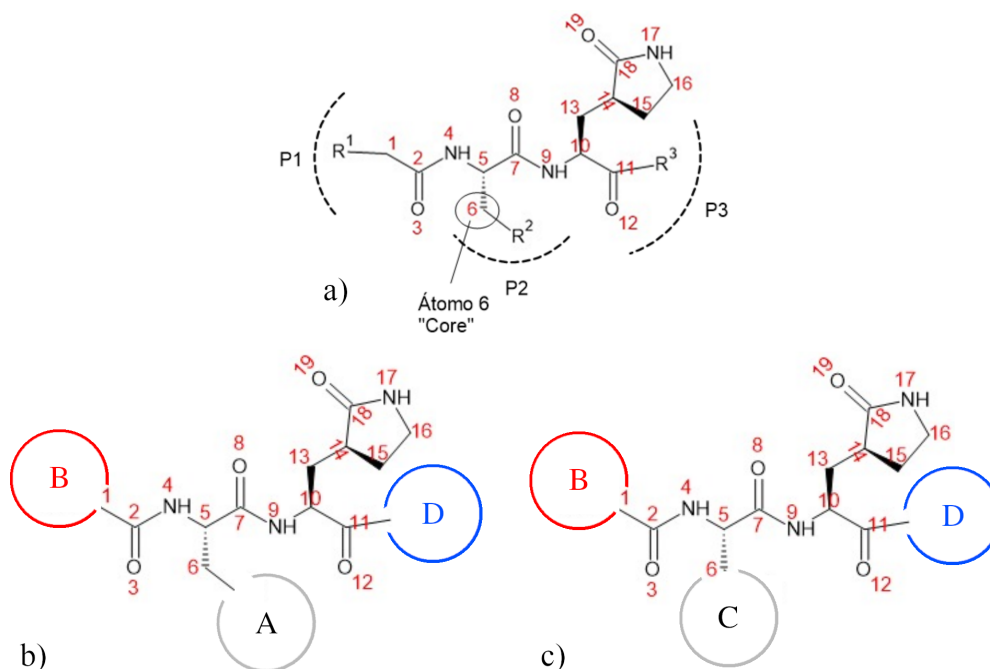
Fonte: Autoria própria

3.1.2. Inibidores α -cetoamidas propostos

Além dos compostos reportados por Dai et al (2020), Zhang et al. (2020a, 2020b) e Yang et al. (2005), novos inibidores derivados de α -cetoamidas foram propostos e avaliados com base nas regras obtidas do estudo de QSAR. A Figura 12 ilustra a estrutura básica comum a todos os compostos, sendo destacadas as regiões de interesse P1, P2 e P3 modificadas e as modificações realizadas. As estruturas propostas foram obtidas por meio da inserção de substituintes laterais, conforme ilustrado nas Tabelas 6-8, mantendo-se a estrutura básica do composto ativo 11a1 ou 11b1. O composto Nirmatrelvir (conhecido comercialmente como o fármaco Paxlovid, desenvolvido pela Pizer) foi também inserido no grupo de testagem, dada a sua similaridade química com os compostos avaliados (incluindo a presença do *core* - vide Figura 13). Desta forma foram definidos os seguintes grupos de testagem:

- Grupo A: substituinte ligado ao átomo 6 (A_1 - A_7);
- Grupo B: substituinte na posição P1, anexado sobre o átomo 1 (B_1 - B_8).
- Grupo C: substituinte na posição P2, anexado sobre o átomo 6 (C_1 - C_9);
- Grupo D: substituinte na posição P3, ligado ao átomo 11 (D_1 - D_{14}).
- Grupo E: Paxlovid.

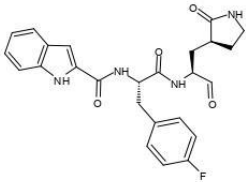
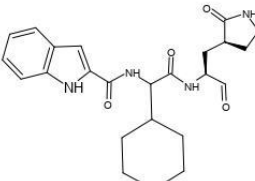
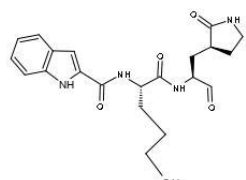
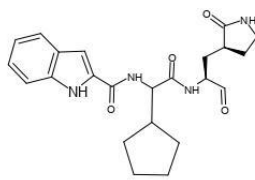
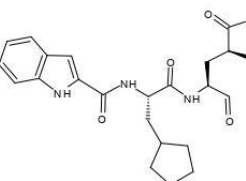
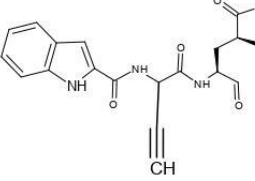
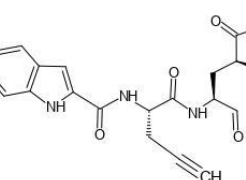
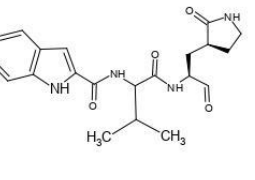
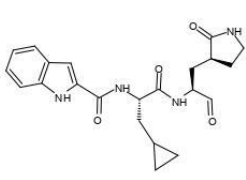
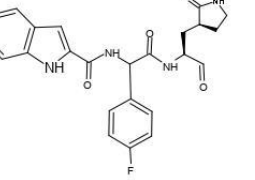
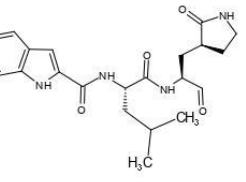
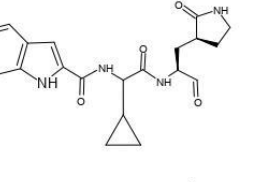
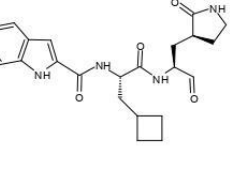
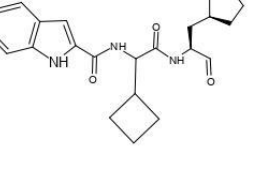
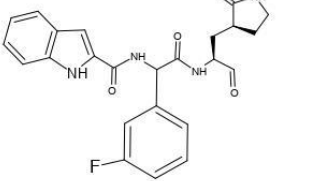
Figura 12. a) Estrutura básica e regiões dos compostos avaliados. b) alterações realizadas sobre os as regiões P1, P2 e P3 para a obtenção dos grupos de testagem A, B e D. c) alterações realizadas sobre os as regiões P1, P2 e P3 para a obtenção dos grupos de testagem B, C e D.



Fonte: Autoria própria.

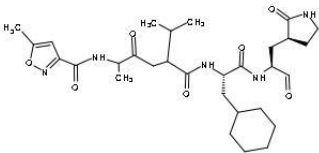
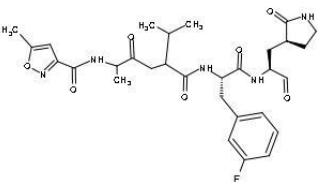
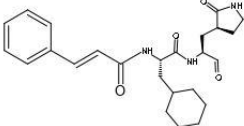
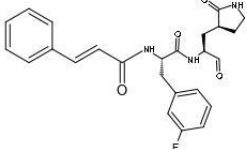
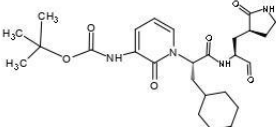
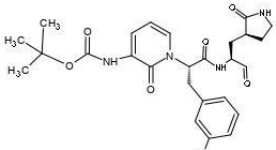
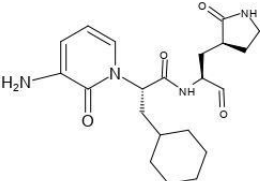
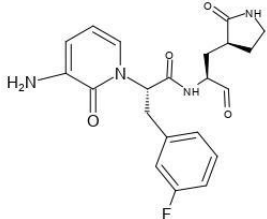
Em especial, as regras de substituição foram planejadas de modo a manter a base dos compostos mais ativos, e substituir apenas uma das regiões: P1, P2 ou P3 para cada derivado.

Tabela 6. Compostos propostos para testagem dos grupos A e C.

composto	Estrutura	composto	Estrutura
A1		C1	
A2		C2	
A3		C3	
A4		C4	
A5		C5	
A6		C6	
A7		C7	
C8		C9	

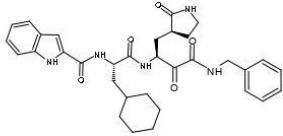
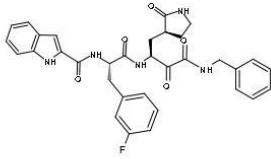
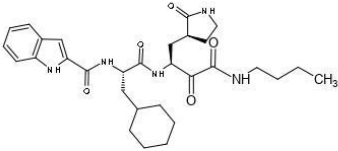
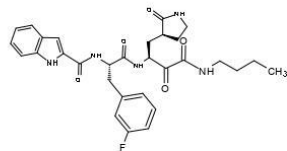
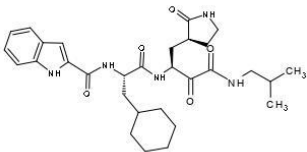
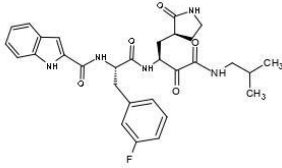
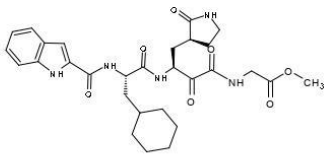
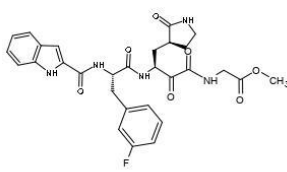
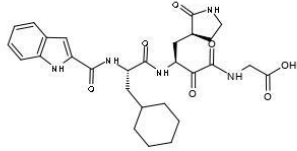
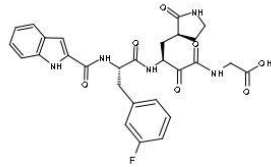
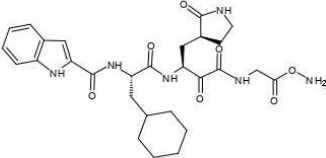
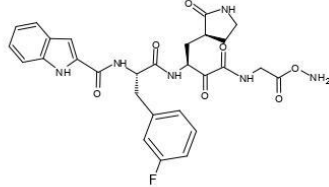
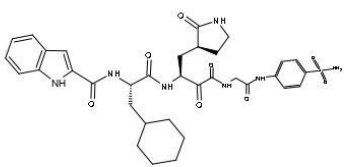
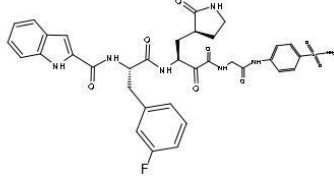
Fonte: Autoria própria

Tabela 7. Compostos propostos para testagem do grupo B.

composto	Estrutura	composto	Estrutura
B1		B2	
B3		B4	
B5		B6	
B7		B8	

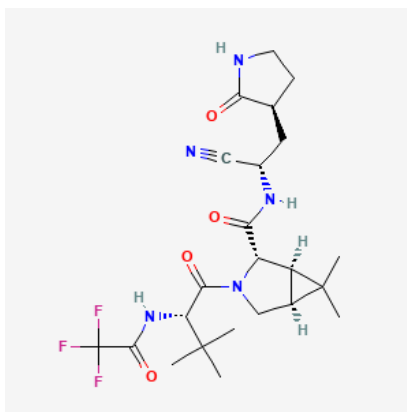
Fonte: Autoria própria.

Tabela 8. Compostos propostos para testagem do grupo D.

composto	Estrutura	composto	Estrutura
D1		D2	
D3		D4	
D5		D6	
D7		D8	
D9		D10	
D11		D12	
D13		D14	

Fonte: Autoria própria

Figura 13. Estrutura bidimensional do Paxlovid grupo de testagem E.



Fonte: National Center for Biotechnology Information (2023)

Uma busca sistematizada envolvendo a combinação de diferentes ligantes laterais poderia levar à criação de um número muito grande de estruturas (665 compostos, para os substituintes considerados no presente trabalho) inviabilizando a presente análise. Optou-se por se investigar mais profundamente a influência do átomo 6 (grupos A e C), dada a sua relevância na regra QSAR obtida.

3.2. Métodos

3.2.1. Estudos conformacionais

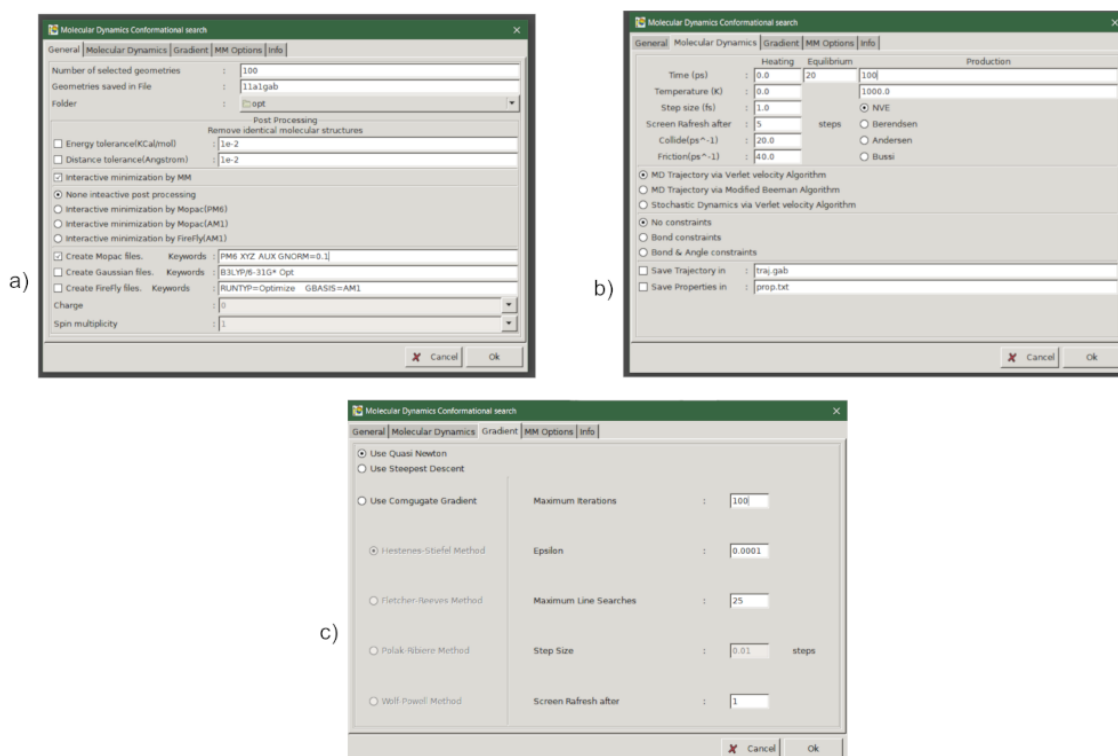
Todas as estruturas foram inicialmente confeccionadas com o auxílio do software ChemSketch (CHEMSKETCH, 2021), e pacotes computacionais Molden (SCHAFTENAAR; NOORDIK, 2000) e Avogadro (HANWELL et al., 2012).

Em especial, por meio do software Chems sketch pode-se, por meio da representação em duas dimensões (2D) dos compostos, obter estruturas 3D, a qual conta com um processo de pré-ajuste da geometria (evitando a sobreposição de átomos). O software Avogadro foi empregado na pré-otimização destas estruturas

iniciais obtidas. Para tanto, foram empregados os campos de forças MMFF94 (HALGREN, 1996) e GAFF (WANG et al., 2004), respectivamente, em moléculas que contém os átomos C, H, O, N, e em moléculas constituídas por C, N, O, H, S, P, F, Cl, Br, e I. Tais campos são amplamente utilizados no estudo de compostos orgânicos e fármacos (TEAM, 2022).

A fim de selecionar conformações de menor energia, estudos de busca conformacional considerando distintas conformações iniciais foram conduzidos por meio de simulações via dinâmica molecular (DM) a altas temperaturas (BATAGIN-NETO et al., 2013a). Tal abordagem visa gerar estruturas com conformação inicial diversificada para posterior otimização e análise via cálculos de estrutura eletrônica, buscando, desta forma avaliar de forma mais apropriada a hipersuperfície potencial associada aos distintos comprimentos e ângulos de ligação, bem como diedrais entre átomos adjacentes dos compostos. Para tanto, os diferentes derivados foram colocados em contato com um reservatório térmico a 1000 K, sendo permitida a sua evolução durante um período de 100 ps após o equilíbrio térmico. A cada 1 ps era coletada uma estrutura para posterior otimização, totalizando 100 estruturas para cada derivado. Os cálculos de DM foram realizados empregando-se o campo de força AMBER (CORNELL et al., 1995) com o auxílio do pacote computacional Gabedit (ALLOUCHE, 2011). Para fins de ilustração, as configurações típicas definidas no software Gabedit para a realização das DM são apresentadas na Figura 14.

Figura 14. Telas de configuração do software Gabedit empregadas no estudo conformacional via DM



Fonte: Autoria própria.

Após a obtenção das estruturas iniciais (100 para cada derivado), as geometrias encontradas foram pré-otimizadas numa abordagem Hartree-Fock e com aproximação semi-empírica PM6 (do inglês, *parametric method 6*) (HF/PM6) (STEWART, 2008), com o auxílio do pacote computacional Mopac2016 (STEWART, 2016) sendo posteriormente identificados confôrmeros de menor energia, por meio da criação de um arquivo formatado de energias para posterior análise gráfica.

Confôrmeros de menor energia global identificados após a pré-otimização via HF/PM6 foram novamente otimizados numa abordagem baseada na teoria do funcional da densidade (DFT), empregando-se o funcional de correlação e troca híbrido B3LYP - Becke, 3 parâmetros, Lee-Yang-Parr (baseado na correção da aproximação local do funcional de correlação e troca pelos termos de troca de HF (1º parâmetro), e Becke 88 (2º parâmetro), e de correlação de Lee, Yang e Parr (3º

parâmetro)). Foram empregadas funções de base *double-zeta* gaussianas polarizadas do tipo 6-31G(d). Tal abordagem tem apresentado resultados relevantes na descrição de propriedades estruturais e eletrônicas de compostos orgânicos em geral (MARDIROSSIAN; HEAD-GORDON, 2017), definindo uma boa relação de custo-benefício para o presente estudo.

Todos os cálculos DFT/B3LYP/6-31G(d) foram conduzidos considerando-se água como solvente implícito por meio de um método contínuo PCM (COSMI et al., 1996). Os cálculos foram realizados com o auxílio do pacote computacional Gaussian 16 (FRISCH, 2016).

3.2.2. Estudo de reatividade

Dados de reatividade local dos sistemas foram avaliados por meio de índices de Fukui condensados aos átomos (IFCA) (YANG; MORTIER, 1986). Tais descritores permitem identificar a posição de sítios moleculares mais propensos para a interação com agentes externos nucleofílicos (f_k^+), eletrofílicos (f_k^-) e/ou radicais livres (f_k^0). Para fins de exemplificação de seus cálculos, considerando um sistema molecular **M** com N elétrons os IFCA's devem ser expressos por:

$$f_k^- = q_k(N) - q_k(N-1) \quad (1)$$

$$f_k^+ = q_k(N+1) - q_k(N) \quad (2)$$

$$f_k^0 = \frac{1}{2} [q_k(N+1) - q_k(N-1)] \quad (3)$$

sendo que $q_k(N+1)$, $q_k(N)$ e $q_k(N-1)$ representam as populações eletrônicas do k -ésimo átomo de **M** considerando suas configurações aniônica (**M**⁻, com $N+1$ elétrons), neutra (**M**, com N elétrons) e catiônica (**M**⁺, com $N-1$ elétrons), obtidos sem a relaxação da estrutura (empregando-se a estrutura otimizada do composto neutro para todas as configurações). Optou-se por se empregar análises populacionais de

Hirshfeld para a obtenção das populações atômicas, uma vez que é conhecido que geralmente tal abordagem resulta em IFCA positivos, de mais fácil interpretação (DE PROFT et al., 2002; ROY; PAL; HIRAO, 1999).

Análises de reatividade adicionais, empregando mapas de potencial eletrostático molecular (PEM) foram também realizadas. Para tanto o PEM foi estimado considerando-se a partição de carga de Chelp (CHIRLIAN; FRANCL, 1987).

Todos os cálculos de reatividade foram conduzidos com o auxílio do pacote computacional Gaussian 16, numa abordagem DFT/B3LYP/6-31G(d).

3.2.3. Obtenção de descritores eletrônicos

A fim de se obter os descritores quânticos a serem empregados no estudo de correlação estatística, cálculos “*single point*” (sem relaxação de geometria) foram realizados para cada uma das estruturas otimizadas. Com base nestes cálculos uma coleção de índices eletrônicos, relacionados com a energia, as ordens de ligação, carga elétrica, dipolo elétrico e nível dos orbitais moleculares de fronteira, potencial de ionização, eletroafinidade, etc, foram obtidos e organizados em um único arquivo de dados para análises estatísticas. Novamente, todos os cálculos foram realizados numa abordagem DFT/B3LYP/6-31G(d), com o auxílio do pacote computacional Gaussian 16. A Tabela 9 apresenta os descritores considerados, divididos em parâmetros globais (para a molécula toda) e locais (que dependem dos átomos considerados).

Tabela 9. Descritores eletrônicos considerados nas análises multivariadas

ID	Sigla	Descrição	Tipo
1	PI	Potencial de ionização	Global
2	AE	Afinidade eletrônica	
3	ω^-	Capacidade eletrodoadora	
4	ω^+	Capacidade eletroaceitora	
5	R_A	Índice de aceitação de elétrons	
6	R_D	Índice de doação eletrônica	
7	V_T	Volume molecular total	
8	A_S	Área molecular superficial total	
9	E_T	Energia Total	
10	E_{HOMO}	Energia do último orbital ocupado (via orbitais de Kohn-Sham)	
11	E_{LUMO}	Energia do primeiro orbital desocupado (via orbitais de Kohn-Sham)	
12-14	D_q	Componente q do momento de dipolo elétrico (q = x, y, ou z)	
15	D_T	Módulo do momento de dipolo total	
16-17	d^{CX-CG}	Distância entre os centróides associados a f^X e o centro geométrico da molécula (X= + ou -)	
18-36	Ch_k	Carga líquida de Hirshfeld associada ao átomo k do core (com k = 1 a 19)	Local
37-74	f_k^X	IFCA associado ao átomo k do core (com k = 1 a 19 e X = + ou -)	

Fonte: Autoria Própria

Os valores de *PI* e *AE* foram estimados por meio de cálculos *single point*, considerando estruturas neutras (com N elétrons) e carregadas (N+1 e N-1 elétrons):

$$PI = E(N - 1) - E(N), \quad (4)$$

$$AE = E(N) - E(N + 1), \quad (5)$$

onde $E(M)$ representa a energia total de um composto com M elétrons (novamente sem relaxação estrutural).

Os parâmetros ω^- e ω^+ foram estimados com base em IP e EA, conforme proposto por Martínez et al. (2008).

$$\omega^- = \frac{(3IP+AE)^2}{16(IP-EA)} \quad (6)$$

$$\omega^+ = \frac{(IP+3AE)^2}{16(IP-EA)} \quad (7)$$

Com base na análise comparativa dos valores de ω^- e ω^+ com descritores de aceitadores e doadores típicos como os átomos de Flúor (F) e o Sódio (Na), pode-se definir os índices R_A e R_D , associados, respectivamente, à capacidade eletroaceitora e eletrodoadora dos compostos, dados por:

$$R_A = \frac{\omega_M^+}{\omega_F^+} \quad (8)$$

$$R_D = \frac{\omega_M^-}{\omega_{Na}^-} \quad (9)$$

Altos valores de R_A e baixos valores de R_D definem moléculas com alta atividade antioxidante (ou seja, com capacidade eletroaceitora e eletrodoadora intensificada). Nas Equações 8 e 9 são tipicamente empregados os valores de $\omega_{Na}^- = 3,46$ e $\omega_F^+ = 3,40$ (MARTÍNEZ et al., 2008).

As posições dos centróides associados a f^* e f ($C(f^*)$ e $C(f)$) foram estimadas considerando-se a média ponderada dos IFCA em relação à posição de cada um dos átomos a eles associados, sendo cada componente q (com $q = x, y$ ou z) dada por.

$$C_q(f^L) = \left(\sum_{k=1}^N f_k^L \cdot q_k \right) / \left(\sum_{k=1}^N f_k^L \right) \quad (10)$$

sendo que q_k representa a componente (x, y ou z) da posição do k -ésimo átomo do composto (contendo N átomos) e f_k^L o valor de IFCA estimado para o k -ésimo átomo (com $L = +$ ou $-$) (note que, por definição, $\sum f_k^L = 1$). Espera-se que as posições de tais

centróides definam centros reativos de relevância de cada molécula, sendo posições espaciais onde se espera facilidade de entrada (f^+) ou saída (f^-) de elétrons. A fim de associar a posição dos centróides em relação aos compostos, calculou-se a distância entre estas coordenadas e o centro geométrico (CG) da molécula para definir o parâmetro $d^{CX-CG} = |r(C(f^+)) - r(CG)|$, o qual é independente da posição da molécula em relação ao sistema de coordenadas considerado.

3.2.4. Análises estatísticas e estabelecimento de correlações e modelos preditivos

Regressões lineares simples (RLS) e múltiplas (RLM) foram primeiramente realizadas com o objetivo de obter equações preditivas capazes de estimar a atividade antiviral dos compostos com base em combinações lineares de alguns descritores eletrônicos. Cuidados foram tomados nesta etapa a fim de evitar efeitos de *overfitting* (sobre-ajuste) das curvas (VENABLES; RIPLEY, 2002), em especial, buscou-se considerar um número reduzido de variáveis independentes (no máximo 3) nas análises de RLM.

Análises de componentes principais (PCA) foram também empregadas a fim de se reconhecer padrões dentre os compostos. Dentro da PCA um novo sistema de coordenadas ortogonal é gerado por combinações lineares dos descritores originais eletrônicos (variáveis independentes), os quais são muito convenientes para a análise de semelhanças entre as moléculas (BROWN, 1988; FERREIRA, 2002). Os eixos principais resultantes, chamados de componentes principais (PCs), são ordenados de acordo com a variação máxima do conjunto de dados. Nesse sentido, o PC_1 apresenta mais informação estatística do que PC_2 ; PC_2 , por sua vez, apresenta mais informação do que PC_3 e assim por diante.

Além dos estudos de relação estrutura atividade (QSAR, do inglês *quantitative structure-activity relationships*) tradicionais, buscou-se aqui desenvolver uma ferramenta 3D de análise de relações estrutura-reatividade-atividade que permitisse a comparação da atividade biológica de compostos com diferentes estruturas básicas (BATAGIN-NETO; LAVARDA, 2013b; NARANJO-MONTOYA et al. 2014; NICOLETI; BATAGIN-NETO; LAVARDA, 2016). Neste sentido, dados estruturais e de reatividade foram projetados em superfícies no sentido de se obter padrões de comparação da atividade dos compostos. Tal abordagem visa não apenas identificar aspectos estruturais relevantes associados à interação receptor-fármaco, mas também identificar processos de transferência de carga (elétrons e prótons) (via IFCA) e interação eletrostática (via mapas de PEM) que ocorrem nestes sistemas (RAMOS-GUZMÁN; RUIZ-PERNÍA; TUÑÓN, 2021). A Figura 15 ilustra este procedimento, com base nas estruturas mais estáveis obtidas do estudo estrutural, foi realizado o alinhamento das moléculas, sendo então criadas superfícies capazes de recobrir todos os sistemas moleculares (com estruturas dissimilares). Os volumes definidos por estas superfícies foram discretizados e as propriedades de interesse (dados de IFCA em PEM) projetadas nos pontos de discretização. O potencial eletrostático molecular ψ_i , associado ao i -ésimo ponto da superfície discretizada foi estimado por:

$$\psi_i = \sum_{k=1}^M \frac{q_k}{|\mathbf{r}_k - \mathbf{r}_i|} \quad (11)$$

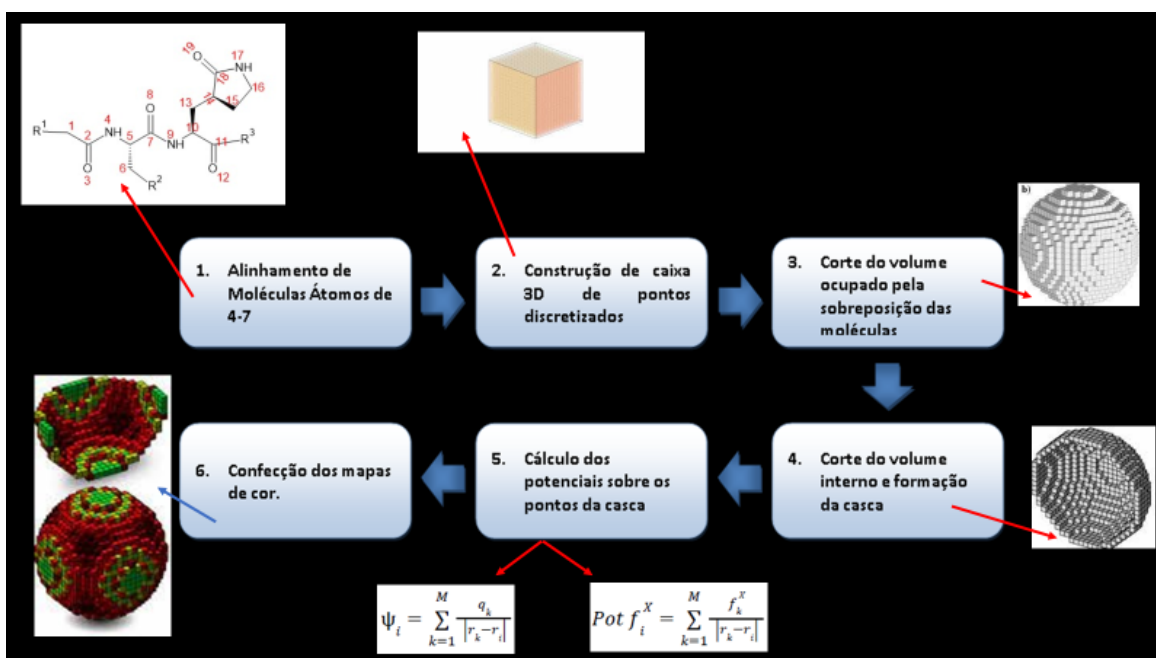
sendo que q_j representa a carga líquida sobre o j -ésimo átomo do composto contendo M átomos (estimada via o método de partição de carga Chelp (CHIRLIAN; FRANCL, 1987)), $\mathbf{r}_i$ e $\mathbf{r}_k$ representam os vetores posição do ponto i da superfície discretizada e do j -ésimo átomo do composto. Valores análogos, definidos como

potenciais de Fukui ($Pot f_i^+$, $Pot f_i^-$ e $Pot f_i^0$), foram estimados considerando os índices f^+ , f^- e f^0 :

$$Pot f_i^X = \sum_{k=1}^M \frac{f_k^X}{|r_k - r_i|} \quad (12)$$

Considerando uma mesma discretização, matrizes tridimensionais com um mesmo número de pontos foram criadas para sistemas com distintas conformações (previamente alinhados), permitindo sua análise comparativa.

Figura 15. Ilustração do processo de construção de superfícies reativas em compostos com estruturas diferentes.



Fonte: Autoria própria

Estudos preliminares envolvendo a seleção de variáveis usando a técnica de aprendizado de máquina baseada em florestas aleatórias (*Random Forests*, RF) foram conduzidos considerando dados de distribuição 3D do potencial elétrico e índices de reatividade. A seleção de variáveis usando RF é um conceito bastante empregado na análise de dados e envolve: i) a criação de uma série de árvores de

decisão binárias aleatórias com diferentes subconjuntos de variáveis do conjunto de dados original; ii) avaliação da importância de cada variável em cada árvore de decisão e com base na melhoria da previsão de propriedades de interesse; iii) cálculo de uma pontuação de importância média para cada variável (*ranking* médio, RM-RF), com base em todas as árvores de decisão. As variáveis RM-RF mais altas são consideradas as mais importantes para a análise e podem ser selecionadas para uso em modelos de previsão ou análises estatísticas (GENUER; POGGI; TULEAU-MALOT, 2010).

O bom desempenho dessa técnica está relacionada, de acordo com Breiman (2001), com a qualidade de cada árvore e com a pequena correlação entre as árvores da floresta, medida pela correlação ordinária das predições das amostras não selecionadas, denominadas *out-of-bag* (OOB). A amostra OOB define um conjunto de observações que não foram usadas para gerar a árvore atual, as quais são empregadas, para se estimar o erro de predição do modelo e avaliar a importância das variáveis.

A medida de importância de cada variável é dada pelo erro quadrado médio do modelo quando os valores observados de uma variável são aleatoriamente permutados nas amostras OOB. Com base neste processo define-se um *ranking* médio (RM-RF) de cada variável dada a sua importância na construção dos modelos nas diferentes repetições. Deste modo, quanto menores os valores de RM-RF (e de seu desvio padrão), mais relevante é a variável nos modelos.

No presente trabalho, as análises de RF foram realizadas em colaboração com o grupo de pesquisa do Prof. Dr. André Debiaso Rossi - UNESP/Itapeva. Visando avaliar características associadas à estrutura 3D, foram consideradas como variáveis originais apenas os potenciais ψ_i e $Pot f_i^X$ (Equações 11 e 12) associados

às malhas formadas. Em especial, dado que o número de estruturas consideradas é muito reduzido (27 moléculas) em relação à dimensão dos dados (5832 variáveis), utilizou-se a técnica de validação cruzada *leave-one-out*. Nesse método, todas as observações são utilizadas para a indução do modelo, exceto uma, que é utilizada para estimar o seu desempenho. Portanto, são gerados 27 modelos (número de casos). A análise para a seleção de variáveis via RF foi realizada com o auxílio do pacote scikit-learn (baseado na classe *RandomForestRegressor*), implementado em Python. Foi considerado um número de árvores igual a 50 (PEDREGOSA et al., 2011).

O erro quadrado médio (MSE, do inglês *mean squared error*) é calculado para as 27 observações. Esse procedimento é repetido 30 vezes, dada a natureza estocástica da RF.

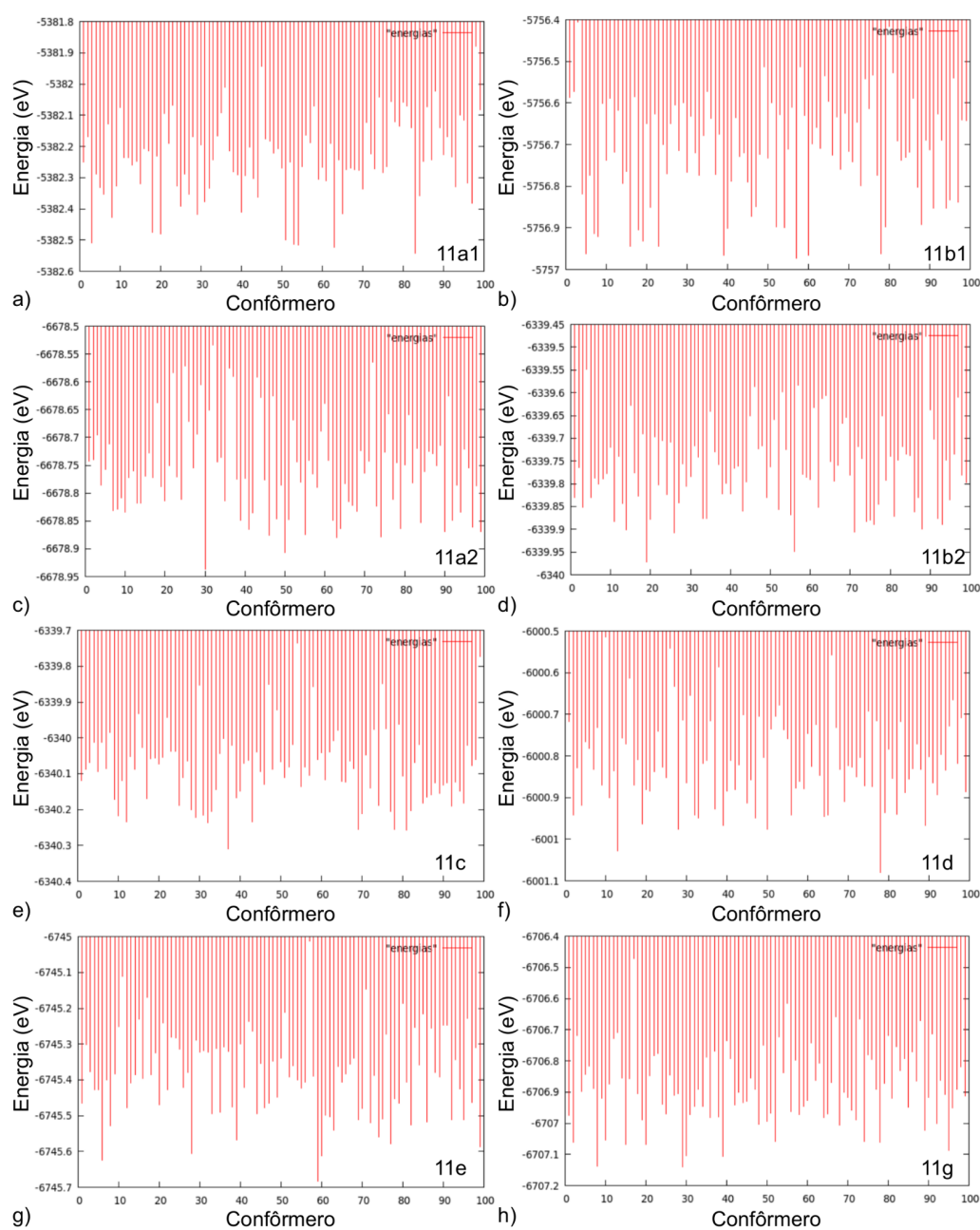
Para saber se o modelo foi de fato capaz de encontrar algum padrão nos dados, comparou-se o MSE dos modelos para as 30 repetições com o MSE calculado a partir da média dos dados de treinamento como preditor da IC₅₀ (método de base). Para o cálculo da importância das variáveis também considerou-se 30 repetições.

4. RESULTADOS

4.1. Estudo estrutural

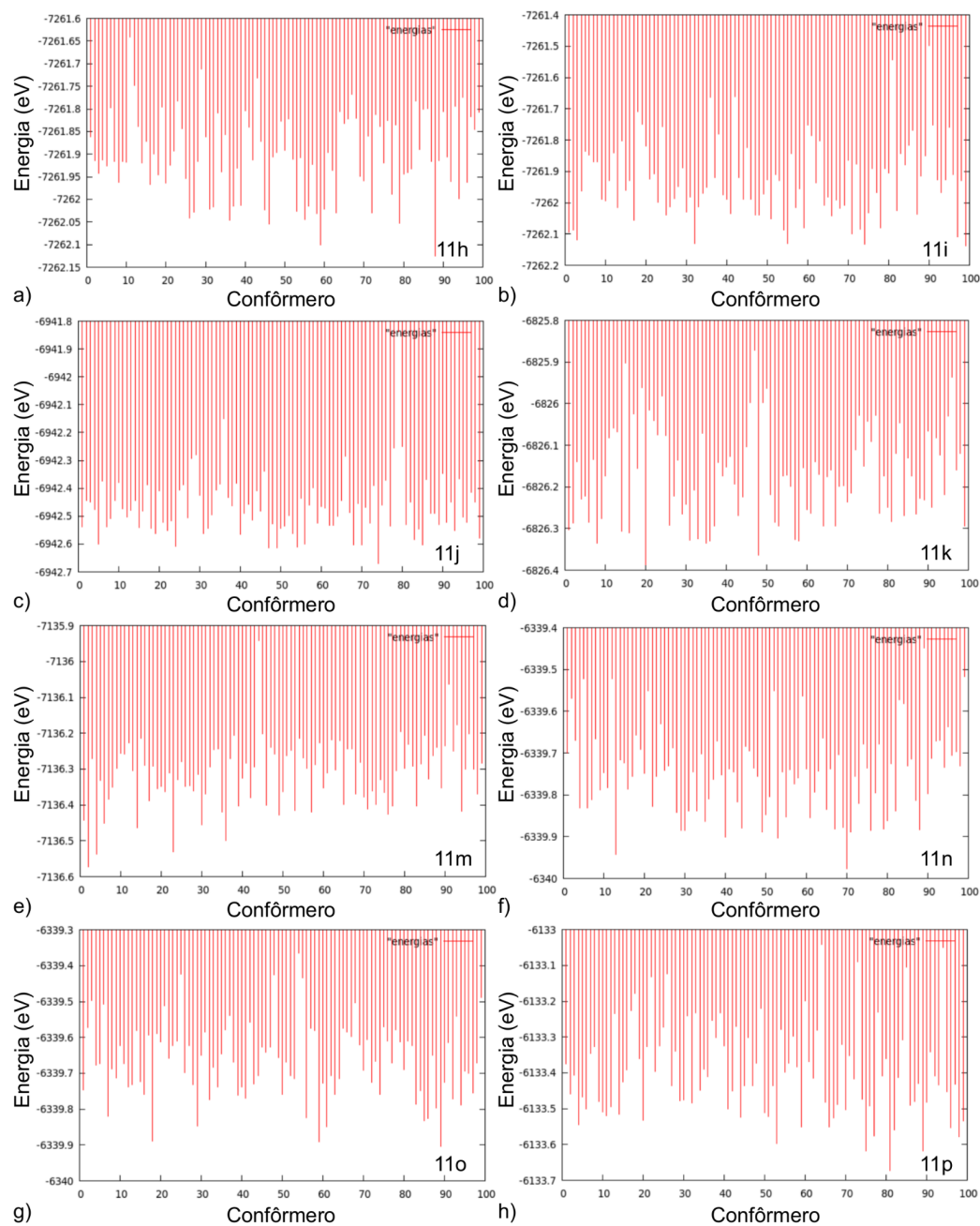
As Figuras 16-19 ilustram as distribuições de energia dos diferentes conformêros de após o processo de pré-otimização via a abordagem HF/PM6.

Figura 16. Distribuições de energia dos conformêros dos compostos a) 11a1, b) 11b1, c) 11a2, d) 11b2, e) 11c, f) 11d, g) 11e, h) 11g obtidas do estudo estrutural.



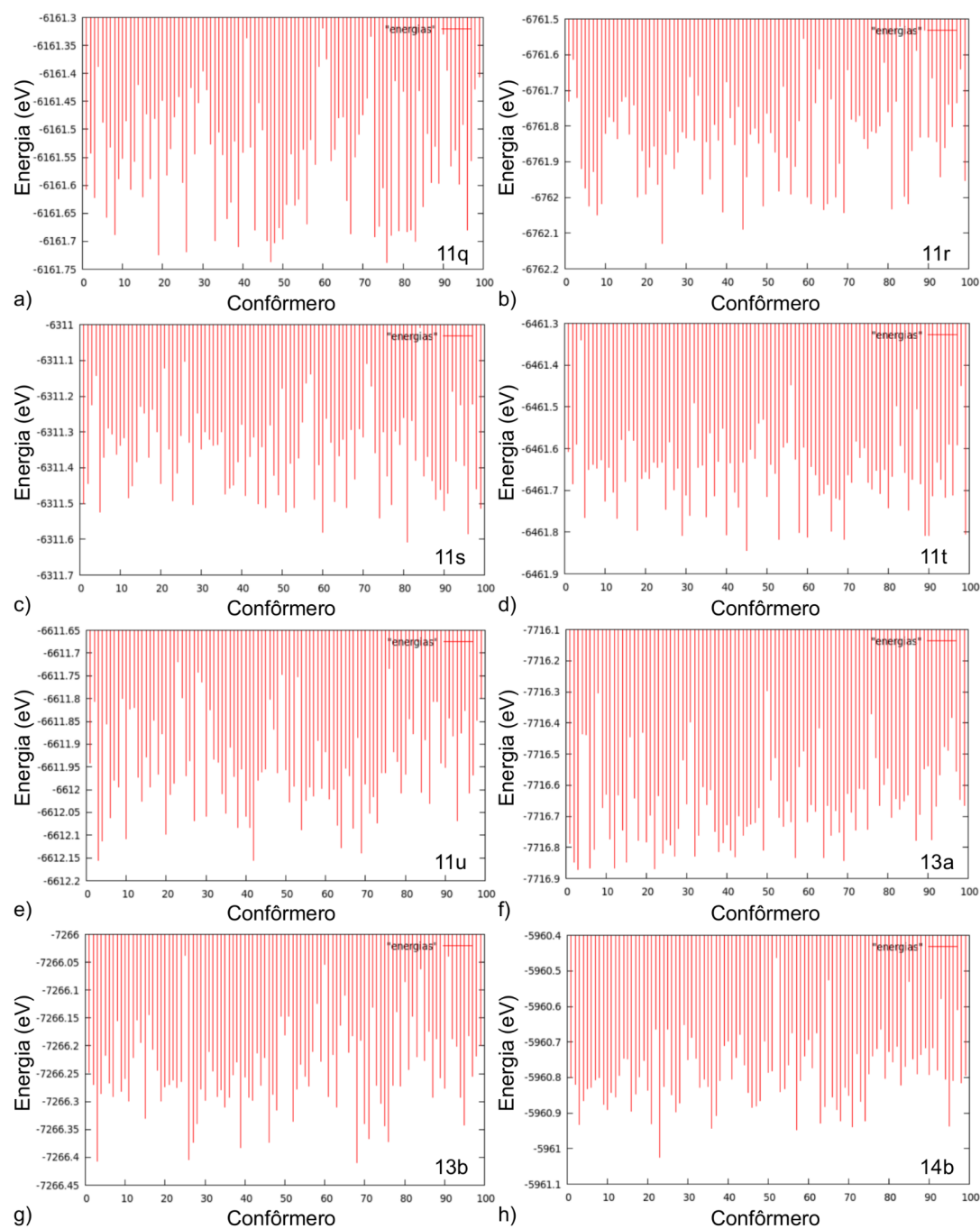
Fonte: Autoria própria

Figura 17. Distribuições de energia dos confôrmers dos compostos a) 11h, b) 11i, c) 11j, d) 11k, e) 11m, f) 11n, g) 11o, h) 11p obtidas do estudo estrutural.



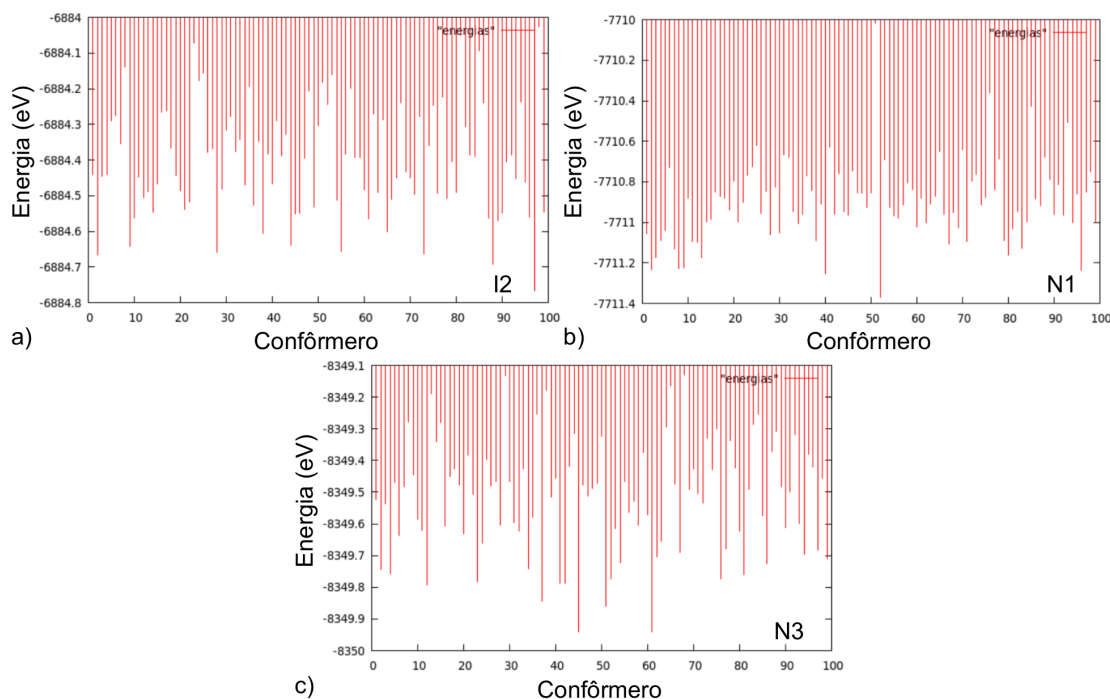
Fonte: Autoria própria

Figura 18. Distribuições de energia dos confômeros dos compostos a) 11q, b) 11r, c) 11s, d) 11t, e) 11u, f) 13a, g) 13b, h) 14a obtidas do estudo estrutural.



Fonte: Autoria própria

Figura 19. Distribuições de energia dos confômeros dos compostos a) I2, b) N1, c) N3, obtidas do estudo estrutural.



Fonte: Autoria própria

De forma geral nota-se que há uma dispersão de energia nos gráficos, tal variação entre a energia de confômeros mais e menos estáveis encontra-se na Tabela 10. Os confômeros que possuem a menor energia são ditos os mais estáveis e são mais prováveis de serem encontrados na natureza, por isso foram considerados nos estudos DFT subsequentes.

Nota-se uma faixa menor de energia em relação aos confômeros 11a2, 11q e 13b, demonstrando que há uma baixa variabilidade estrutural nestes derivados. Por outro lado, o composto N1 apresenta uma faixa de variação de energia maior que os demais compostos, indicando maior flexibilidade estrutural deste derivado.

Tabela 10. Faixa de energia da diferença entre o conformero mais estável e menos estável.

Composto	Energia dos conformeros		Variação de energia (eV)	Composto	Energia dos conformeros		Variação de energia (eV)
	Menos estável (eV)	Mais estável (eV)			Menos estável (eV)	Mais estável (eV)	
11a1	-5381,878	-5382,543	0,665	11o	-6339,364	-6339,903	0,538
11a2	-6678,534	-6678,936	0,402	11p	-6133,042	-6133,672	0,630
11b1	-5756,406	-5756,972	0,566	11q	-6161,319	-6161,738	0,418
11b2	-6339,476	-6339,972	0,496	11r	-6761,521	-6762,129	0,608
11c	-6339,735	-6340,309	0,574	11s	-6311,009	-6311,607	0,598
11d	-6000,515	-6001,080	0,565	11t	-6461,340	-6461,844	0,504
11e	-6745,013	-6745,682	0,670	11u	-6611,674	-6612,155	0,481
11g	-6706,473	-6707,141	0,668	13a	-7716,134	-7716,872	0,738
11h	-7261,642	-7262,124	0,483	13b	-7266,038	-7266,410	0,372
11i	-7261,498	-7262,137	0,640	14b	-5960,464	-5961,024	0,560
11j	-6941,837	-6942,670	0,834	I2	-6884,027	-6884,765	0,739
11k	-6825,871	-6826,387	0,515	N1	-7710,014	-7711,372	1,357
11m	-7135,943	-7136,572	0,630	N3	-8349,129	-8349,941	0,812
11n	-6339,449	-6339,976	0,528				

Fonte: Autoria própria

A Tabela 11 ilustra quais conformeros foram selecionados a partir das análises dos gráficos de energia para a condução dos estudos DFT.

Tabela 11. Confôrmers selecionados de menor energia.

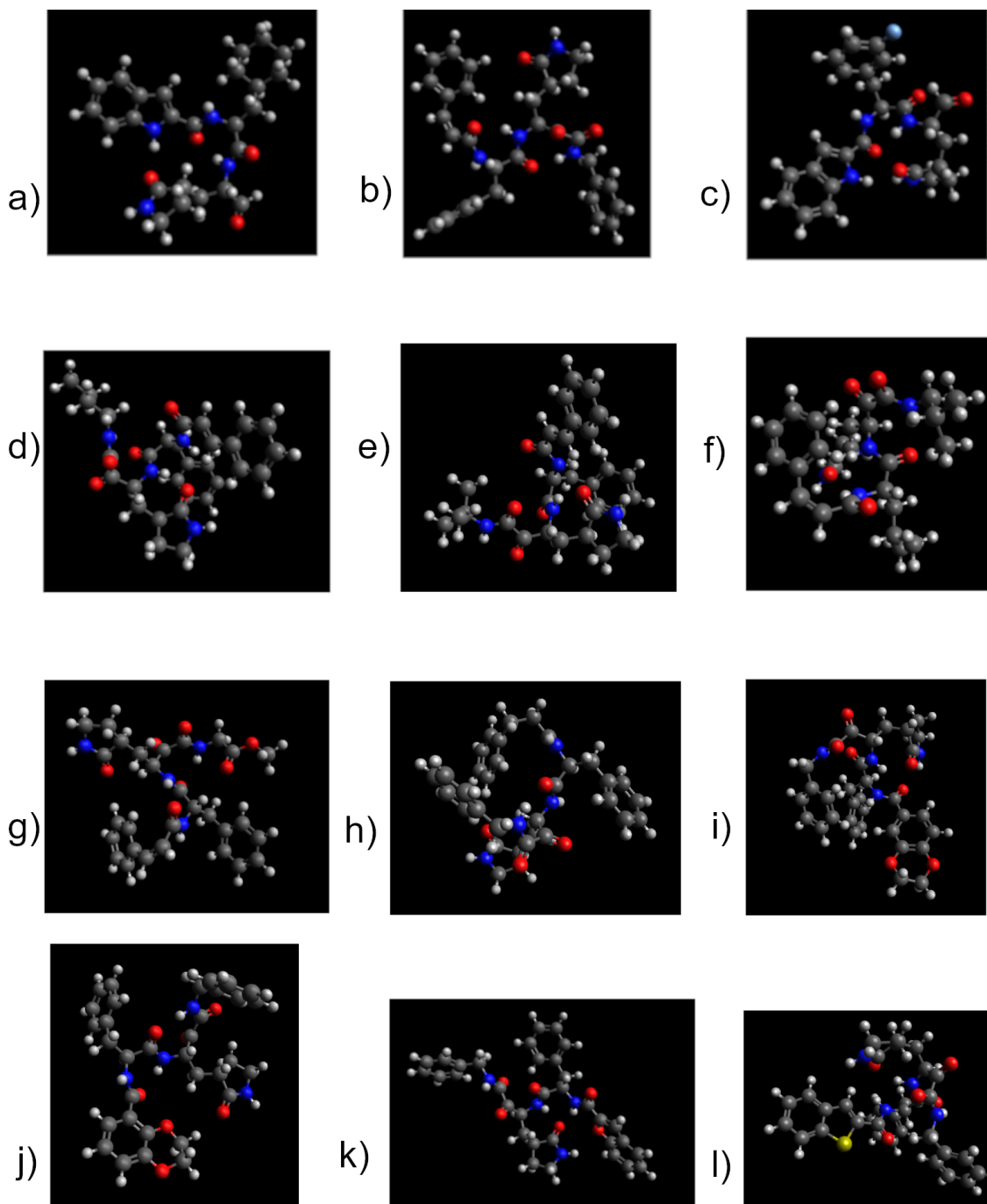
Composto	Confôrmero Selecionado	Composto	Confôrmero Selecionado	Composto	Confôrmero Selecionado
11a1	83	11i	99	11s	81
11b1	57	11j	74	11t	45
11a2	30	11k	20	11u	42
11b2	19	11m	2	13 ^a	3
11c	37	11n	70	13b	68
11d	78	11o	89	14b	23
11e	59	11p	81	l2	97
11g	29	11q	76	N1	52
11h	88	11r	24	N3	45

Fonte: Autoria própria

As estruturas acima apresentadas foram então completamente re-otimizadas numa abordagem DFT/B3LYP/6-31G(d). As Figuras 20-22 ilustram as conformações de menor energia obtidas deste processo de otimização.

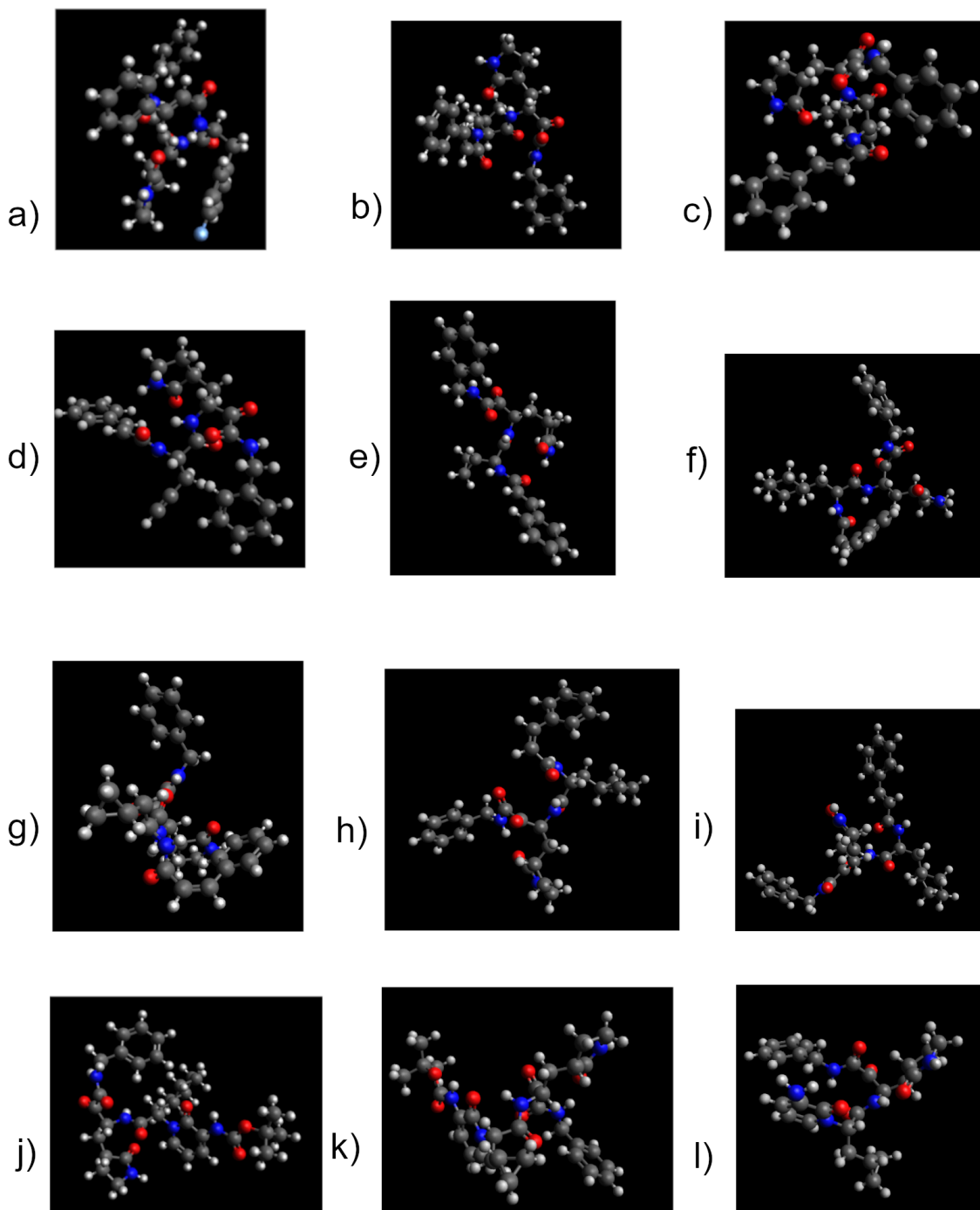
Nota-se na maioria dos casos que as estruturas tendem dispor seus anéis em posições relativamente distantes do *core* principal, sugerindo que tais estruturas podem desempenhar um papel relevante na atividade biológica destes sistemas. As estruturas otimizadas obtidas foram utilizadas nos estudos de reatividade subsequentes.

Figura 20. Estruturas de menor energia obtidas do estudo estrutural após otimização via abordagem DFT/B3LYP/6-31G(d), sendo: a) 11a1 b) 11a2 c) 11b1 d) 11b2 e) 11c f) 11d g) 11e h) 11g i) 11h j) 11i k) 11j l) 11k



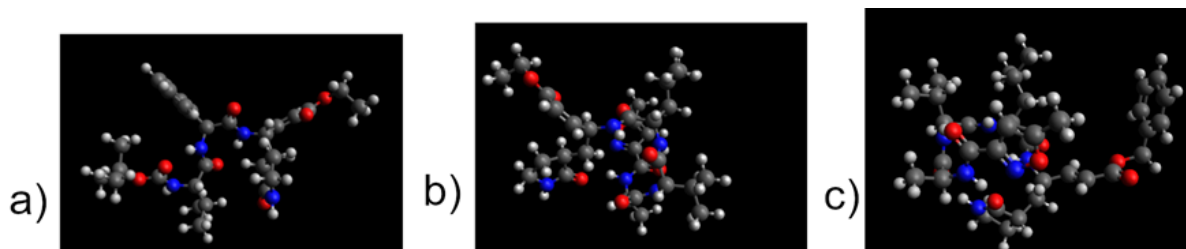
Fonte: Autoria própria

Figura 21. Estruturas de menor energia obtidas do estudo estrutural após otimização via abordagem DFT/B3LYP/6-31G(d), sendo a) 11m b) 11n c) 11o d) 11p e) 11q f) 11r g) 11s h) 11t i) 11u j) 13a k) 13b l) 14b



Fonte: Autoria própria

Figura 22. Estruturas de menor energia obtidas do estudo estrutural após otimização via abordagem DFT/B3LYP/6-31G(d), sendo a) I2 b) N1 c) N3.



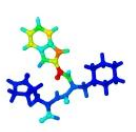
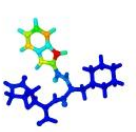
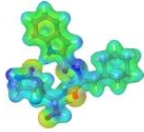
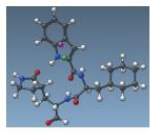
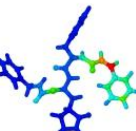
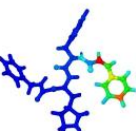
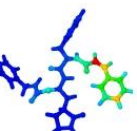
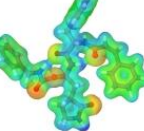
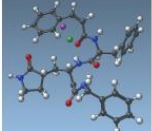
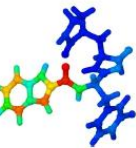
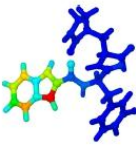
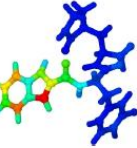
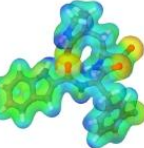
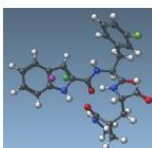
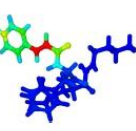
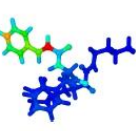
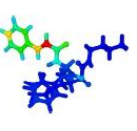
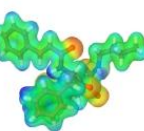
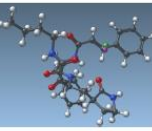
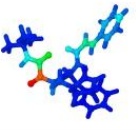
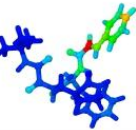
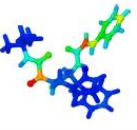
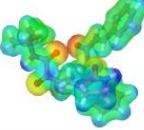
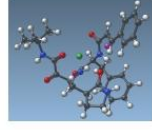
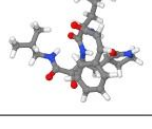
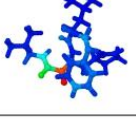
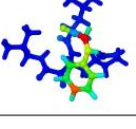
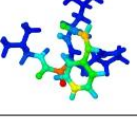
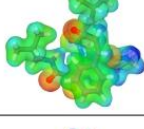
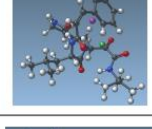
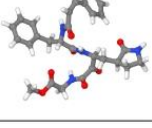
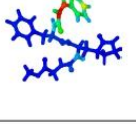
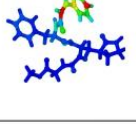
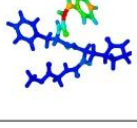
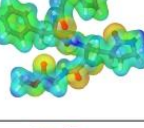
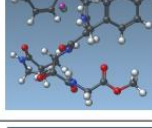
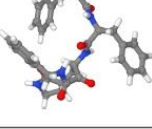
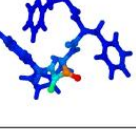
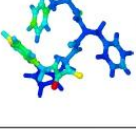
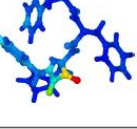
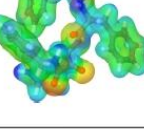
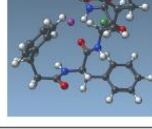
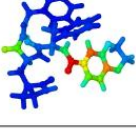
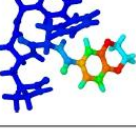
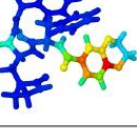
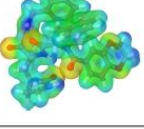
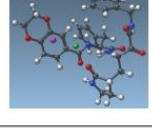
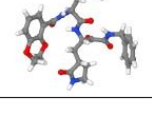
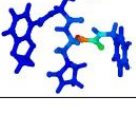
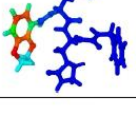
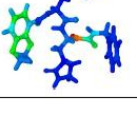
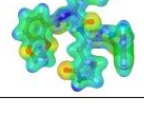
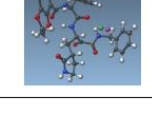
Fonte: Autoria própria

4.2. Estudo de reatividade

A fim de avaliar aspectos de reatividade dos sistemas estudados baseados em IFCA e potencial eletrostático molecular foram realizados, no sentido de se identificar regiões susceptíveis a interações com agentes eletrofílicos (f), nucleofílicos (f^+) e radicais livres (f^0) em cada derivado.

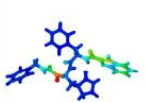
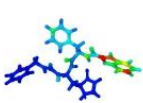
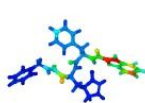
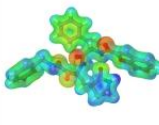
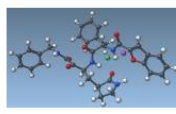
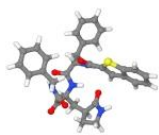
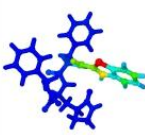
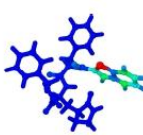
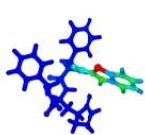
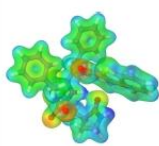
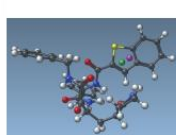
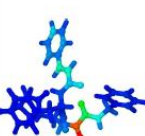
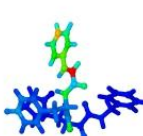
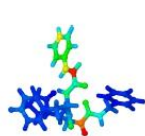
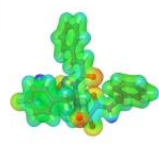
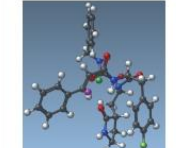
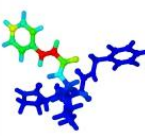
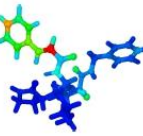
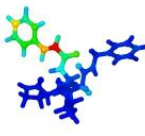
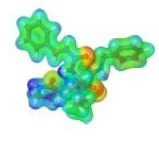
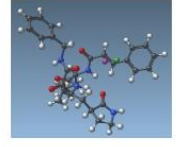
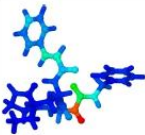
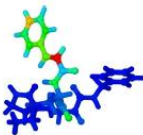
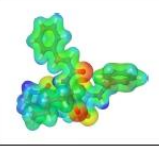
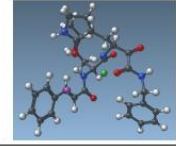
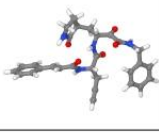
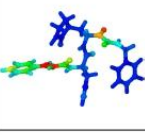
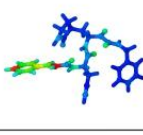
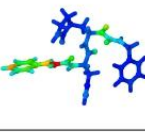
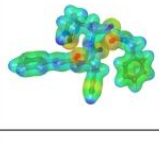
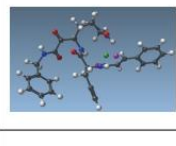
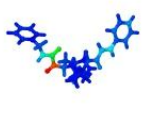
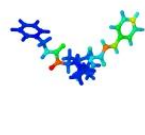
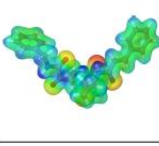
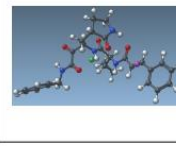
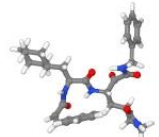
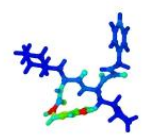
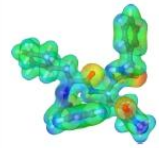
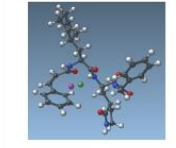
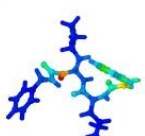
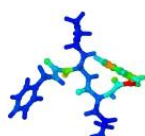
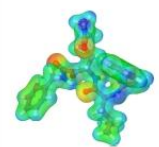
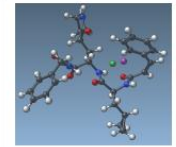
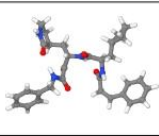
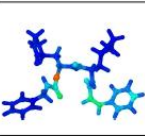
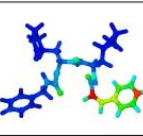
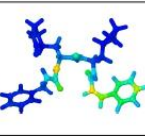
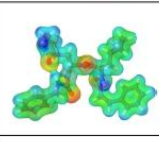
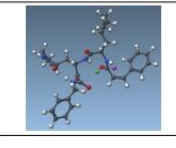
As Figuras 23-25 ilustram os mapas de cor associados aos Índices de Fukui Condensados ao Átomos (IFCA). Regiões em vermelho e azul representam sítios moleculares reativos e inertes, respectivamente. As demais cores indicam sítios com reatividade intermediária, seguindo uma escala RGB (vermelho > laranja > amarelo > verde > azul). Também são apresentadas as posições de centróides associados à f^+ (em verde) e f (em magenta) de forma tridimensional, obtidos considerando-se a média ponderada dos IFCA em relação à posição de cada um dos átomos a eles associados. Espera-se que as posições de tais centróides definam centros reativos de relevância de cada molécula, sendo posições espaciais onde se espera facilidade de entrada (f^+) e saída (f) de elétrons. Projeções bidimensionais desses centróides são apresentadas no Apêndice 1.

Figura 23. IFCA, PEM e Centr ides dos inibidores: 11a1, 11a2, 11b1, 11b2, 11c, 11d, 11e, 11g, 11h e 11i.

Composto	Estrutura	f^+	f^-	f^0	PEM	Centroids 3D
11a1						
11a2						
11b1						
11b2						
11c						
11d						
11e						
11g						
11h						
11i						

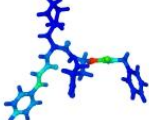
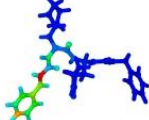
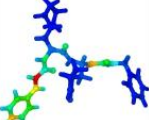
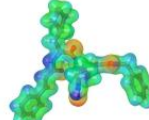
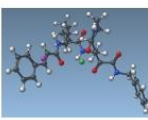
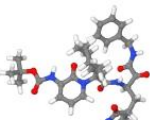
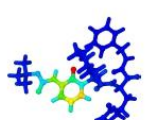
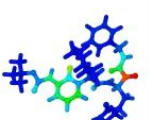
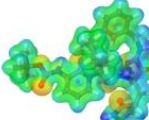
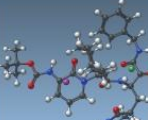
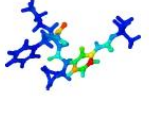
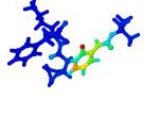
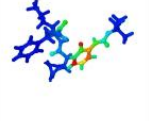
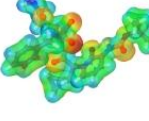
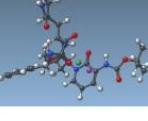
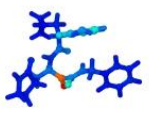
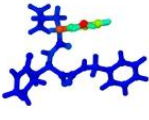
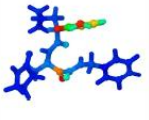
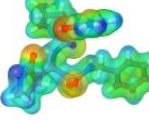
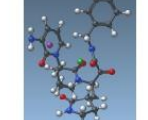
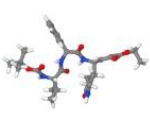
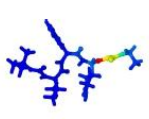
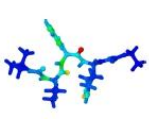
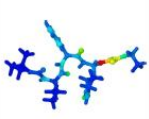
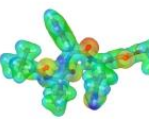
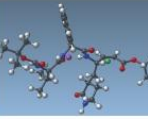
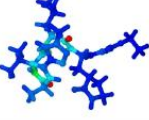
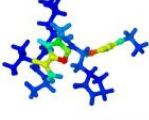
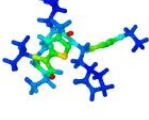
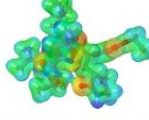
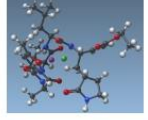
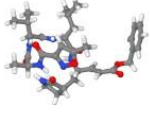
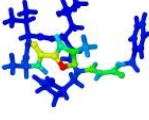
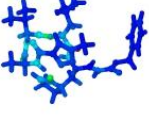
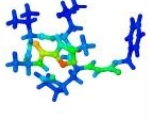
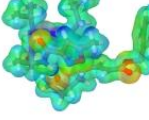
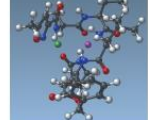
Fonte: Autoria pr pria

Figura 24. IFCA, PEM e Centr ides dos inibidores: 11j, 11k, 11m, 11n, 11o, 11p, 11q, 11r, 11s e 11t.

Composto	Estrutura	f^+	f^-	f^0	PEM	Centroids 3D
11j						
11k						
11m						
11n						
11o						
11p						
11q						
11r						
11s						
11t						

Fonte: Autoria pr pria

Figura 25. IFCA, PEM e Centróides dos inibidores: 11u, 13a, 13b, 14b, l2, N1 e N3.

Composto	Estrutura	f^+	f^-	f^0	PEM	Centroids 3D
11u						
13a						
13b						
14b						
l2						
N1						
N3						

Fonte: Autoria própria

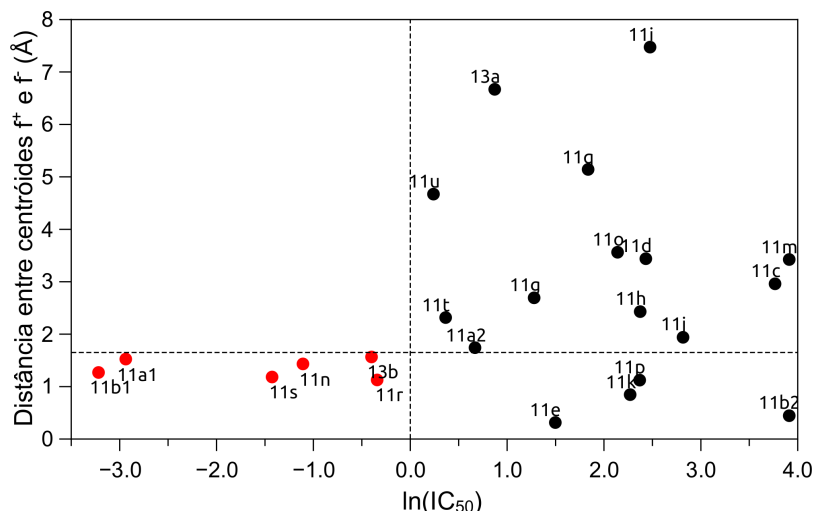
De forma geral, com relação aos IFCA, nota-se que R^1 e R^3 possuem a maior relevância para a reatividade dos diferentes derivados de α -cetoamidas, sendo notada uma maior reatividade sobre estes sítios na maioria dos compostos estudados. Em geral os compostos mais ativos (com $IC_{50} < 1,0$, ou seja 11b1, 11a1, 11s, 11n, 13b e 11r) apresentam maior reatividade nos anéis associados à R^1 .

Com relação ao PEM nota-se uma concentração de cargas positivas sobre os átomos de nitrogênio e hidrogênio dos derivados, enquanto que uma carga líquida negativa é observada sobre os átomos de oxigênio. Não há padrões que permitam diferenciar compostos ativos de inativos associados ao PEM.

De forma geral nota-se os centros reativos magenta (associados a f^+) encontram-se próximos a regiões mais negativamente carregadas, como o oxigênio e átomos de carbonos com insaturações, enquanto que centróides associados a f^- encontram-se mais próximos a regiões com deficiência de elétrons (em geral carbonos das cetonas). É interessante notar que, os derivados mais ativos (com menores valores de IC_{50}) apresentam, como característica em comum, a presença de ambos os centróides nas imediações do sítios associados a R^1 , sugerindo que estes sítios desempenham um papel relevante na atividade inibitória destes compostos.

A Figura 26 ilustra como a distância $d^{C^+-C^-} = |r(C(f^+)) - r(C(f^-))|$ entre os centróides permite diferenciar alguns dos compostos mais ativos (em vermelho) dos demais (em preto). Em geral, derivados com $d_{cutoff} < 1,65 \text{ \AA}$ definem compostos com $\ln(IC_{50}) < 1,0$, contudo 4 exceções são também identificadas: 11b2, 11e, 11k e 11p, o que sugere que outros parâmetros estruturais, não considerados na Figura 26, devem desempenhar um papel relevante na atividade antiviral dos compostos, sendo esta uma condição necessária, porém não suficiente para definir sistemas bioativos. Tal característica sugere que possíveis processos de troca eletrônica (retirada inserção de elétrons) relevantes no processo de inibição ocorram em regiões geometricamente próximas (possivelmente sobre os mesmos sítios, interagindo com aminoácidos específicos da M_{pro})

Figura 26. Influência da distância, d , entre centróides de f^+ e f^- na classificação de compostos com atividade antiviral (círculos vermelhos definem compostos ativos $\ln(IC_{50}) < 0.0$, enquanto que círculos em preto indicam compostos de menor atividade antiviral)



Fonte: Autoria própria

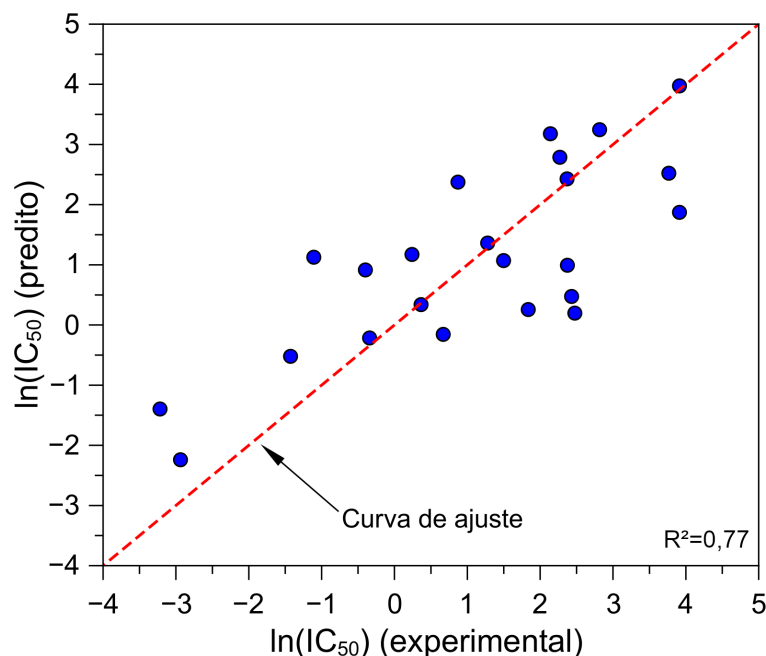
Visando identificar tais características, estudos de QSAR foram avaliados para identificação de padrões de correlação.

4.3. Estudos de QSAR

A Figura 27 ilustra dados previstos via RLM que apresentaram a melhor correlação estatística entre a estrutura eletrônica e atividades antivirais dos compostos (com coeficiente de determinação $R^2 = 0,77$). Tal relação foi obtida considerando-se o $\ln(IC_{50})$ como variável dependente (regressões menos expressivas foram obtidas ao se considerar os valores de IC_{50} e $1/IC_{50}$ como variáveis dependentes). A Equação 13 apresenta os descritores empregados na RLM, sendo que: E_T representa a energia total do composto; Ch_5 a carga de Hirshfeld estimada sobre o átomo 5 da estrutura básica dos compostos e f_6^- indica o índice de reatividade relativo a espécies eletrofílicas sobre o átomo 6 da estrutura básica (a numeração considerada nos átomos é apresentada na Figura 10).

$$\ln(\text{IC}_{50}) = 3,832 + (-4,328 \times 10^{-3}) \cdot E_T + (-3,723 \times 10^2) \cdot \text{Ch}_5 + (1,787 \times 10^2) \cdot f_6^- \quad (13)$$

Figura 27. Predição da atividade biológica em escala logarítmica nos eixos x e y.



Fonte: Autoria própria

Tal resultado permite estimar as seguintes características como desejáveis em derivados com maior atividade antiviral: i) altos valores de E_T , o que pode ser associado a estrutura com um maior número de elétrons, ou seja, ligantes maiores em geral com sistemas ressonantes; ii) carga líquida positiva sobre o átomo 5; e iii) baixos valores de f_6^- , o que sugere uma baixa reatividade sobre o átomo 6 em relação a agentes eletrofílicos externos.

Tais descritores, de fato, estão associados à distribuição espacial dos centros reativos identificada na seção anterior, sugerindo que uma baixa reatividade na região central dos derivados é de fato desejável a fim de facilitar a interação e troca de elétrons com o ambiente externo. Além disso, os dados apontam a relevância de

uma estrutura rica em elétrons, sugerindo que sistemas ressonantes na região dos ligantes é de fato, uma característica de interesse em compostos ativos.

Nenhuma discriminação relevante entre compostos ativos e inativos pode ser identificada via análises de PCA, sendo os resultados omitidos por simplicidade.

4.4. Proposição e prospecção de novos compostos

Dados os resultados advindos do estudo de reatividade e QSAR, diferentes derivados foram propostos. Para tanto, um planejamento molecular de 38 inibidores foi realizado tomando-se como base a estrutura de dois inibidores originais de maior atividade: 11a1 e 11b1, considerando-se as regiões de interesse P1, P2 (A6P2) e P3, além do composto comercial Paxlovid (vide Figura 12 e Tabelas 6-8). Estudos estruturais similares aos realizados para os compostos originais foram conduzidos para os derivados dos 5 grupos de planejamento, sendo seus valores de $\ln(IC_{50})$ estimados com base na Equação 13.

As Tabelas 12-16 e Figura 28 sumarizam os dados de predição da atividade biológica dos diferentes grupos de testagem. A faixa em cinza representa a região sobre a qual se distribuem os valores de $\ln(IC_{50})$ dos compostos reportados na literatura (ou seja, $-3,21 < \ln(IC_{50}) < 3,91$). Em especial, nota-se compostos com valores previstos de IC_{50} inferiores a 1 μ M ($\ln(IC_{50}) < 0$).

Tabela 12. Atividade antiviral prevista para os compostos do grupo de testagem “A”.

Grupo A	ln(IC₅₀)	IC₅₀ (μM)
1	-1,068	0,344
2	-2,498	0,082
3	-1,577	0,207
4	-4,180	0,015
5	-2,221	0,109
6	-1,776	0,169
7	-2,009	0,134

Fonte: Autoria própria.

Tabela 13. Atividade antiviral prevista para os compostos do grupo de testagem “B”.

Grupo B	ln(IC₅₀)	IC₅₀ (μM)
1	-0,174	0,840
2	30,116	>50
3	-0,917	0,400
4	0,526	1,693
5	0,612	1,843
6	0,811	2,249
7	-1,636	0,195
8	28,025	>50

Fonte: Autoria própria.

Tabela 14. Atividade antiviral prevista para os compostos do grupo de testagem “C”.

Grupo C	ln(IC₅₀)	IC₅₀ (μM)
1	-0,930	0,394
2	0,572	1,772
3	-7,184	0,001
4	0,208	1,231
5	-0,614	0,541
6	-2,867	0,057
7	-2,333	0,097
8	-0,987	0,373
9	-0,370	0,691

Fonte: Autoria própria.

Tabela 15. Atividade antiviral prevista para os compostos do grupo de testagem “D”.

Grupo D	ln(IC₅₀)	IC₅₀ (μM)	Grupo D	ln(IC₅₀)	IC₅₀ (μM)
1	2,037	7,666	8	2,150	8,587
2	2,154	8,619	9	0,454	1,574
3	1,700	5,475	10	1,793	6,009
4	-2,024	0,132	11	0,462	1,587
5	1,102	3,010	12	0,828	2,288
6	-1,838	0,159	13	3,576	35,728
7	0,351	1,421	14	2,619	13,720

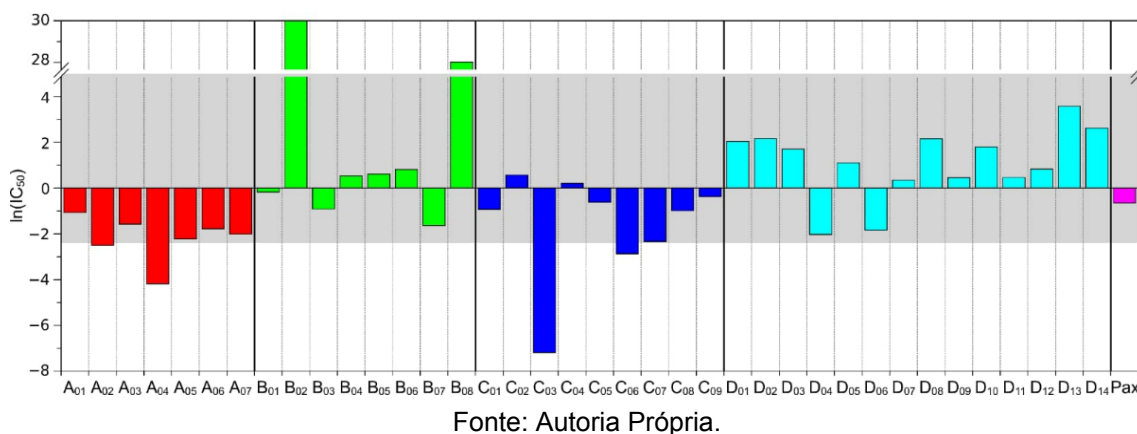
Fonte: Autoria própria.

Tabela 16. Atividade antiviral prevista para o Paxlovid, desenvolvido pela Pfizer.

Grupo E	ln(IC₅₀)	IC₅₀ (μM)
Paxlovid	-0,648	0,523

Fonte: Autoria própria.

Figura 28. Predição da atividade biológica em escala logarítmica dos novos sistemas moleculares investigados.



Nota-se que os grupos de testagem A e C são os mais promissores, apresentando um grande número de estruturas com valores de IC₅₀ inferiores a 1. Em especial deve-se destacar os compostos C₀₃, A₀₄, C₀₆, A₀₂ e C₀₇ com valores de IC₅₀ inferiores a 0,1 µM. Deve-se destacar que todos os compostos do grupo A apresentam valores previstos de IC₅₀ < 1 µM ($\ln(IC_{50}) < 0$), enquanto que duas exceções são observadas no grupo C (C₀₂ e C₀₄)

Por outro lado, nota-se valores superiores a 1 µM ($\ln(IC_{50}) > 0$) para grande parte dos compostos pertencentes aos grupos B e D, com exceção dos compostos B₀₁, B₀₃, B₀₇, D₀₄ e D₀₆, os quais apresentam valores relativamente menores de IC₅₀.

De modo geral os dados acima apresentados sugerem que a presença do grupo indol na posição P1 da estrutura básica das α-cetoamidas (Figura 12) é uma característica desejável a inibidores ativos. Tal proposição é reforçada pela alta atividade observada para o grupo C (o qual também apresenta este grupo em P1) e atividade reduzida no grupo B, onde os substituintes de P1 são alterados, levando, geralmente, a performances inferiores.

Outro dado relevante é o papel fundamental do substituinte anexado em P3 (vide Figura 12). Como se pode notar, os dados sugerem que a presença de ligantes

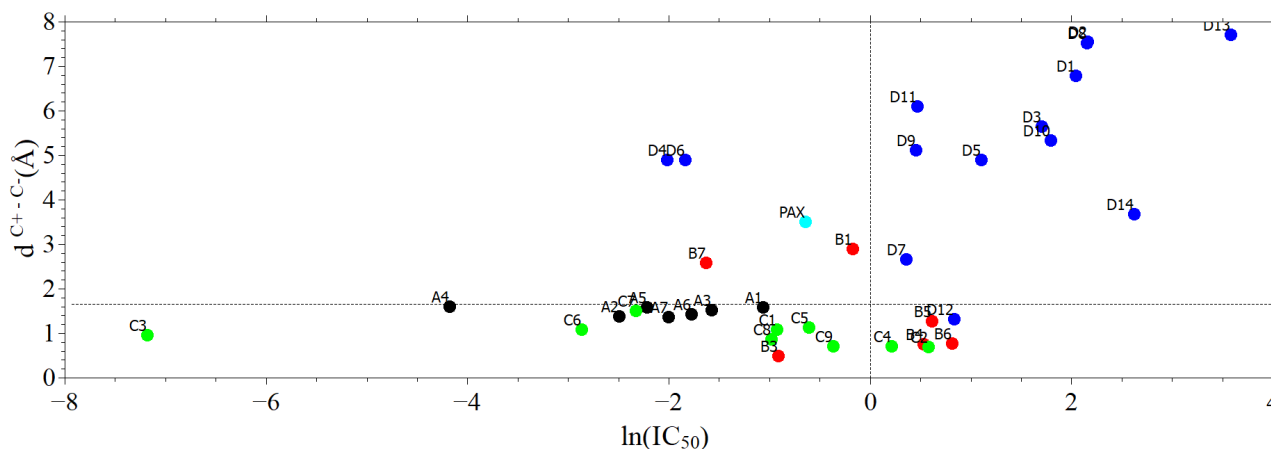
nesta região leva a uma redução da atividade biológica, como notado no grupo de testagem D. Tal resultado é compatível com a redução da interação da enzima M_{pro} com a região P3, considerada essencial no mecanismo de inibição covalente das α -cetoamidas, conforme proposto por Ramos-Guzmán, Ruiz-Pernía e Tuñón (2021). Deste modo, apesar da Equação 13 (QSAR) não conter de forma explícita dados relativos aos átomos da posição 11 e 12 na atividade dos compostos, o seu efeito indireto acaba sendo identificado em certo grau.

Por fim é interessante mencionar a capacidade da Equação 13 em prever o composto Paxlovid como um derivado ativo, com IC_{50} inferior a $0,5 \mu M$, compatível com dados experimentais que sugerem valores de IC_{50} ao redor de $0,02 \mu M$ com relação ao SARS-CoV 2. (LI et al., 2022; PAXLOVID, 2022)

Visando avaliar a relevância da distância inter-centróides na identificação de compostos promissores, identificada previamente para o conjunto de compostos reportados na literatura (Figura 26), buscou-se avaliar estes dados para os distintos grupos de testagem. A Figura 29 ilustra os resultados obtidos. Os dados relativos à distribuição espacial dos IFCA's, e representação das posições dos centróides dos grupos de testagem são apresentados nos Apêndices III e IV, por simplicidade.

Note que, de forma geral, compostos ativos (com $\ln(IC_{50}) < 0$) seguem a proposta de apresentarem distâncias inter-centróides inferiores a $1,65 \text{ \AA}$. Em especial, todos os compostos A_n e C_n seguem esta regra, inclusive o composto ativo B_{03} . Temos como *outliers* os compostos D4 e D6, B1 e B7, e PAX, como compostos ativos que apresentam distâncias maiores que as esperadas. Vale a pena também mencionar que os 5 compostos de maior atividade predita (C_{03} , A_{04} , C_{06} , A_{02} e C_{07}) seguem este padrão.

Figura 29. Influência da distância, d , entre centróides de f^+ e f^- na classificação de compostos com atividade antiviral dos grupos de testagem (círculos vermelhos definem compostos ativos $\ln(IC_{50}) < 0.0$, enquanto que círculos em preto indicam compostos de menor atividade antiviral)



Fonte: Autoria própria

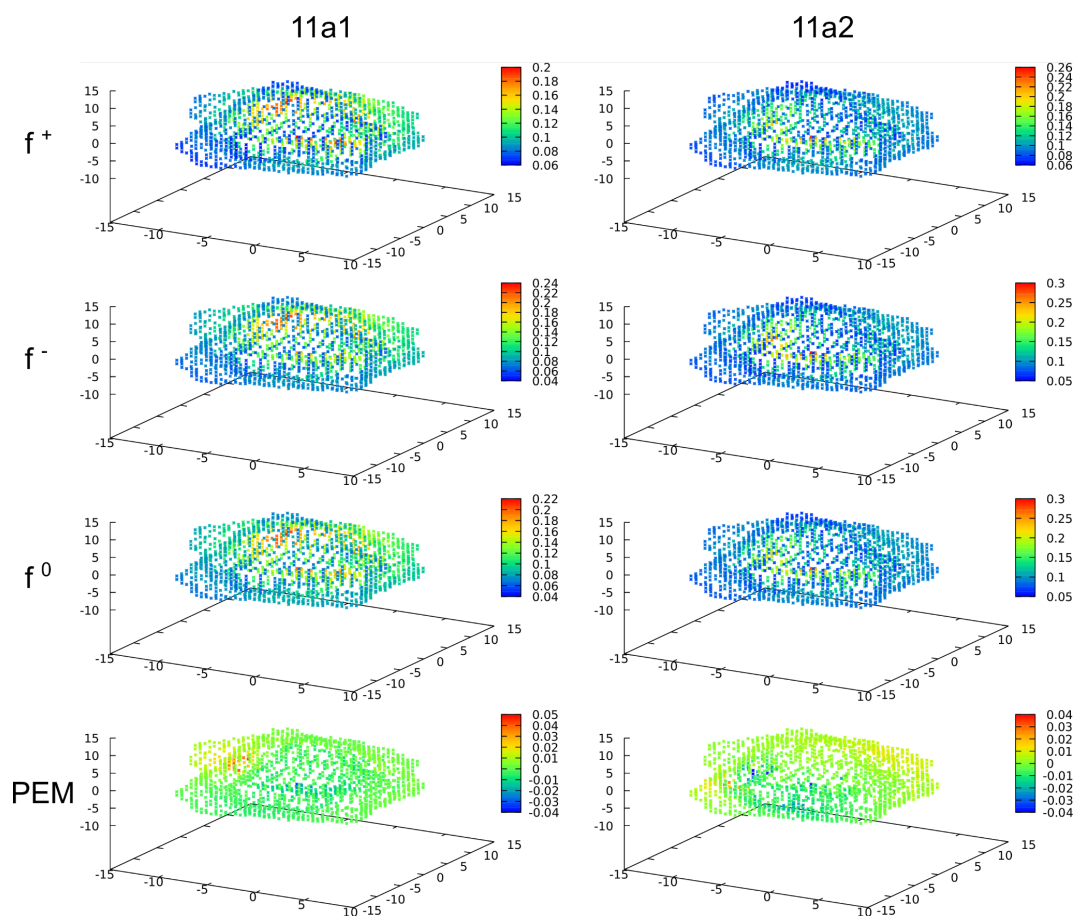
De forma geral os resultados sugerem que a presença de uma estrutura ressonante, como o indol, em P1 favorece uma alta atividade dos compostos, ao mesmo tempo em que a anexação de estruturas em P3 reduz a propriedade de interesse dos derivados de α -cetoamidas. A influência de um grupo ressonante em P1 pode também explicar parcialmente os menores valores de IC_{50} obtidos para B1, B3 e B7 dentro do grupo de testagem B, composto principalmente por derivados inativos. O papel do substituinte anexado em P2, apesar de relevante (dada a Equação 13) não segue um padrão definido (estruturas ressoantes, não ressoantes etc), contudo a sua escolha pode modular os parâmetros Ch_5 e f_6^- , identificados como relevantes no estudo QSAR. Em geral uma escolha apropriada do grupo anexado em P3 deve levar a uma intensificação da carga positiva (ou redução da carga negativa) sobre o átomo 5 do *core* e/ou redução da reatividade sobre o átomo 6 do *core* em relação a agentes externos eletrofílicos.

4.5. Análise 3D de descritores de reatividade

A análise tridimensional também foi realizada em relação aos descritores de reatividade IFCA e PEM. Para tanto foi construída uma superfície associada ao descritor PEM e três superfícies associadas aos IFCA. Essas superfícies foram construídas a partir de uma distribuição de pontos capaz de englobar todos os 27 inibidores da literatura (previamente alinhados). Inicialmente as estruturas dos inibidores foram alinhadas com base nos átomos 4, 5, 6 e 7 do *core*, construiu-se então uma caixa com 15 Å de aresta com pontos discretizados (de 1 em 1 Å), sobre os quais foram estimados os valores de potencial eletrostático e “potenciais de Fukui”. Tal caixa foi reduzida a uma casca capaz de englobar as 27 estruturas alinhadas considerando-se as distâncias com os átomos mais externos. Tais cálculos foram realizados com o auxílio de um programa especificamente desenvolvido em linguagem Fortran 90 juntamente com o orientador.

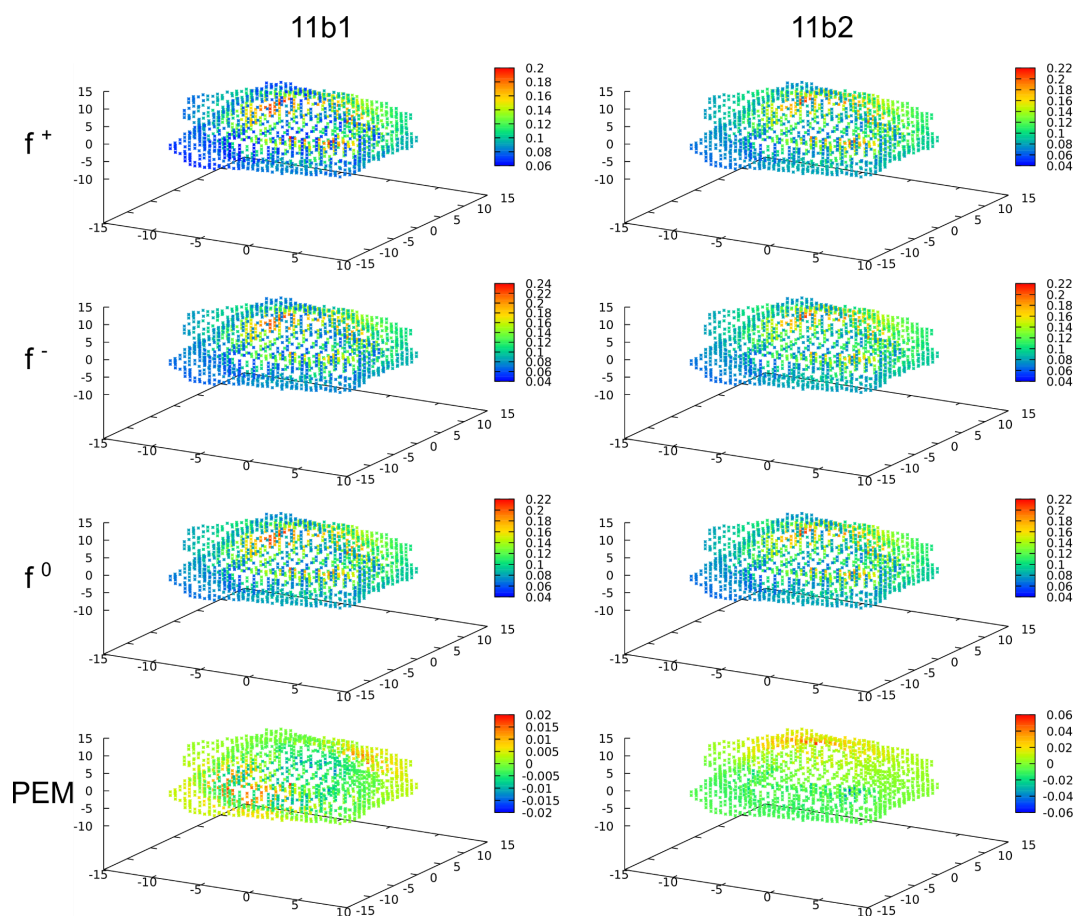
As Figuras 30-32 ilustram as superfícies dos compostos: 11a1 e 11a2; 11b1 e 11b2, também 11m e 11n, respectivamente, compostos de maior ($IC_{50} < 1,0 \mu M$; 11a1, 11b1 e 11n) e menor ($IC_{50} > 50 \mu M$; 11b2 e 11m) atividades antivirais (11a2 apresenta uma atividade intermediária com $IC_{50} = 1,950 \mu M$). As representações associadas às demais estruturas são apresentadas no Apêndice II.

Figura 30. Representação das Superfícies Moleculares 3D com os descritores IFCA e PEM: 11a1 e 11a2. Os eixos coordenados x, y e z, representam as coordenadas espaciais em Å (omitidas por simplicidade).



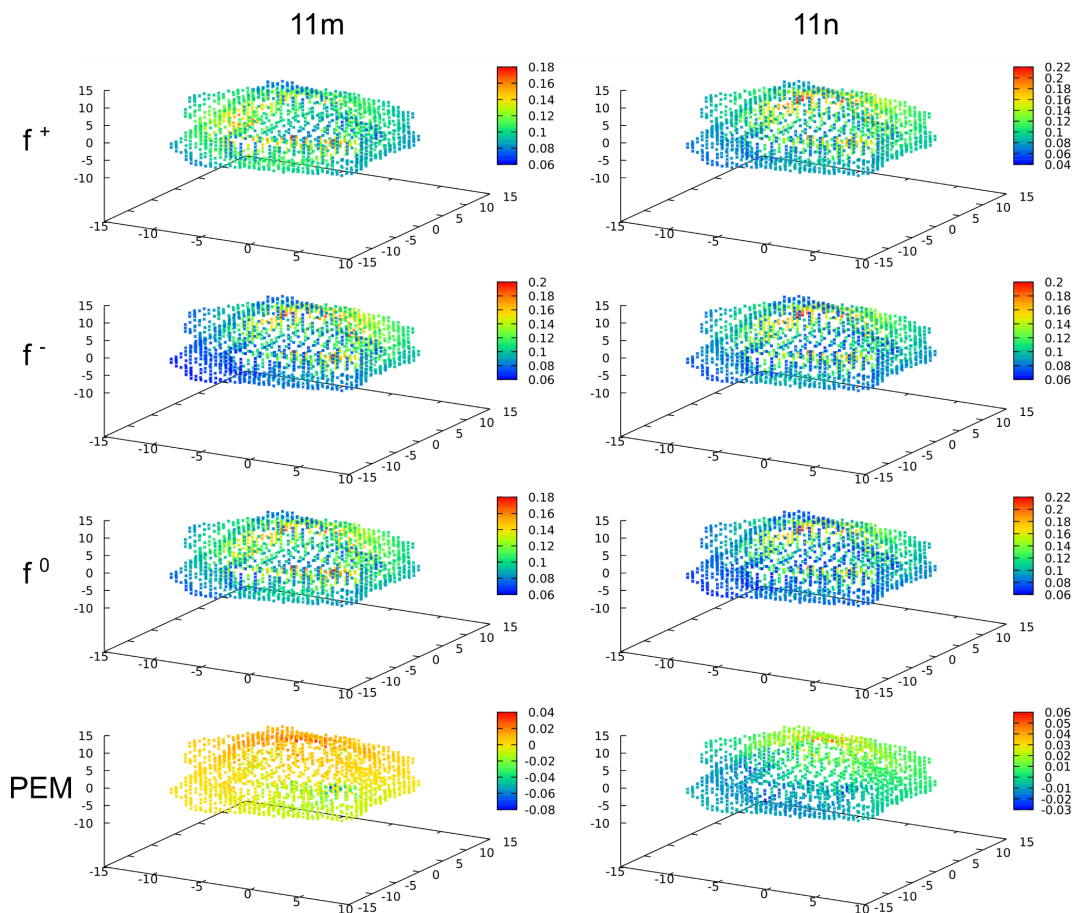
Fonte: Autoria própria.

Figura 31. Representação das Superfícies Moleculares 3D com os descritores IFCA e PEM: 11b1 e 11b2. Os eixos coordenados x, y e z, representam as coordenadas espaciais em Å (omitidas por simplicidade).



Fonte: Autoria própria.

Figura 32. Representação das Superfícies Moleculares 3D com os descritores IFCA e PEM: 11m e 11n. Os eixos coordenados x, y e z, representam as coordenadas espaciais em Å (omitidas por simplicidade).



Fonte: Autoria própria.

Uma análise preliminar dos resultados sugere que compostos que apresentam valores de potencial eletrostático muito positivo sobre a parte superior das superfícies moleculares 3D. Apresentam menor atividade, como pode ser notado no caso da estrutura 11m e 11b2. Os compostos mais ativos apresentam regiões mais negativamente carregadas nos equadores das superfícies, enquanto que regiões mais positivas situam-se nas proximidades de regiões mais reativas definidas pelos IFCA.

Visando identificar padrões tridimensionais e características desejáveis em compostos antivirais eficientes, estudos preliminares envolvendo técnicas de

aprendizado de máquina e seleção de variáveis usando RF foram empregadas. Foram consideradas como variáveis originais para a construção dos modelos os valores dos potenciais ψ_i e $Pot f_i^X$ (Equações 11 e 12) associados aos 1458 pontos das malhas formadas após o alinhamento das estruturas. Totalizou-se assim 5832 variáveis de *input* no modelo (1+3 conjuntos associados a ψ_i , $Pot f_i^+$, $Pot f_i^-$, $Pot f_i^0$).

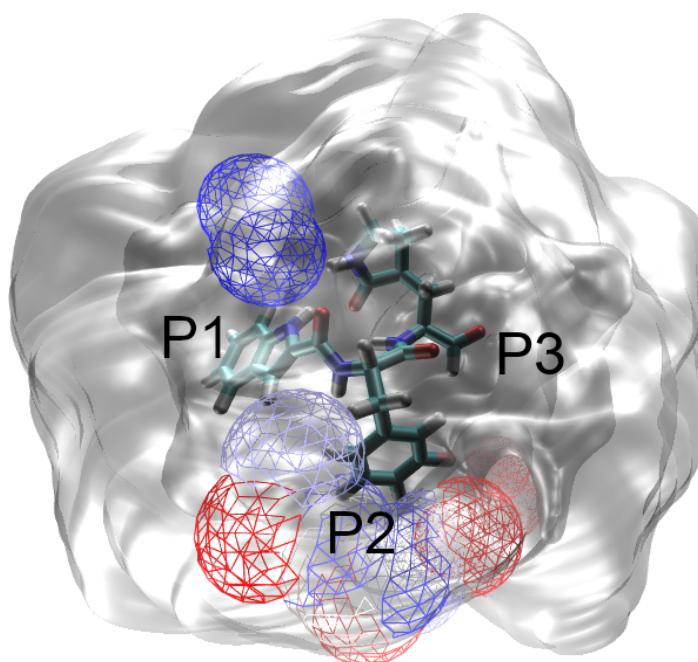
Um valor de MSE=3,9678 foi obtido pelo método base (análise de médias), enquanto que por meio de RF obteve-se MSE=3,0970 com desvio padrão de 0,2124. Isso significa que as RF geradas nas 30 repetições foram capazes de encontrar algum padrão nos dados. Dessa maneira, de forma geral, pode-se considerar que a importância das variáveis atribuída pelos modelos RF foram úteis na obtenção de um modelo preditor do problema analisado. As variáveis de maior importância foram selecionadas considerando-se os menores RM-RF (de maior prioridade nas diferentes repetições). Deste modo, foram tomadas as variáveis com RM-RF < 200 e com os menores desvios padrão, totalizando 17 descritores.

A Figura 33 ilustra a malha de pontos comum aos fármacos (superfície translúcida) enquanto que os círculos coloridos em formato de malha indicam a posição dos descritores mais relevantes identificados nos modelos de RF, todos eles associados ao potencial eletrostático. Tais pontos são apresentados seguindo uma escala RGB de intensidade (vermelho descritores de mais alta prioridade e azul os de prioridade reduzida).

É interessante salientar que todos os descritores considerados como os de maior prioridade estão associados à distribuição do potencial eletrostático sobre a superfície comum dos compostos. Neste contexto, a ausência dos descritores associados aos “potenciais de Fukui” evidenciam que interação dos fármacos com espécies químicas externas é dominada por interações eletrostáticas em detrimento

de interações envolvendo os orbitais moleculares (previstas pelos IFCAs). É importante notar que a maior parte dos sítios de relevância situam-se ao redor da região P2, em concordância com a Eq. 13. Em especial a distribuição espacial dos descritores é aproximadamente perpendicular à linha que liga as regiões P1 e P3, indicando a relevância de interações estéricas adicionais. Tais estudos serão aprofundados em etapas futuras do trabalho.

Figura 33. Representação das Superfícies Moleculares 3D com a identificação de sítios relevantes identificados via análises de *Random Forest*. O composto 11b1 é ilustrado para fins de identificação das regiões P1, P2 e P3.



Fonte: Autoria própria.

5. CONCLUSÕES

No presente trabalho buscou-se estabelecer relações entre a estrutura eletrônica de uma família de α -cetoamidas e a atividade antiviral destes compostos contra os vírus SARS-CoV e SARS-CoV 2. Buscou-se assim identificar características eletrônicas, estruturais e de reatividade relevantes na atividade antiviral de compostos tidos como promissores no tratamento da COVID-19 e desenvolver uma metodologia de análise de relações estrutura-propriedade empregando-se índices de reatividade e análise de estruturas 3D.

Estudos de IFCA indicam que os sítios mais reativos das α -cetoamidas situam-se preferencialmente sobre R^1 e R^3 . Em particular nota-se que a distância entre os centróides (C) associados a centros aceitadores (f^+) e doadores (f^-) define um parâmetro de relevância na atividade biológica destes compostos. Em especial, uma distância entre centróides inferior à 1,65 Å define uma condição necessária (observada em todos os compostos ativos) porém não suficiente (uma vez que há exceções) na identificação de sistemas antivirais de alta performance.

Estudos de QSAR indicam a relevância do substituinte R^2 , na atividade dos compostos por meio da carga líquida sobre o átomo 5 (Ch_5) e o IFCA associado a interações com eletrófilos sobre o sítio 6 (f_6^-). Em particular, baixas reatividades sobre a região central dos derivados é desejável a fim de facilitar a interação e troca de elétrons com o ambiente externo (possivelmente por meio de R^1). Uma regra simples de predição de atividade foi identificada, a qual permite selecionar compostos mais promissores para a condução de ensaios experimentais.

Com base nestes dados, novos derivados foram propostos e avaliados (grupos de testagem A, B, C, D e E). Os resultados sugerem a relevância da

presença de estruturas ressonantes sobre P1 e ausência de ligantes em P3 como uma característica desejável a compostos ativos. Por outro lado, nota-se que os substituintes anexados em P2 tem um papel relevante na modulação da carga líquida e reatividade sobre regiões centrais dos derivados.

Estudos preliminares de potenciais moleculares 3D foram conduzidos no sentido de se identificar padrões desejáveis à compostos ativos via técnicas de aprendizado de máquina. Os resultados indicam a relevância da região P2 na atividade antiviral destes compostos, reforçando os dados de QSAR.

Em suma, os resultados obtidos do presente trabalho ao mesmo tempo em que permitem estabelecer uma regra simples para proposição de compostos antivirais promissores baseados em α -cetoamidas, propõe uma metodologia diferenciada baseada em cálculos DFT e técnicas de aprendizado de máquina para identificação de sítios relevantes na ação inibitória desses compostos que podem ser empregada em outros sistemas.

REFERÊNCIAS

ALLOUCHE, A. R. Gabedit-A graphical user interface for computational chemistry softwares. **Journal of Computational Chemistry**, v. 32, n. 1, p. 174–182, 2011.

BALMANT, O.; FONTANETTO, R. **Haverá mais pandemias depois da covid-19? Provavelmente sim**. 2022. Disponível em: <https://www.terra.com.br/vida-e-estilo/saude/havera-mais-pandemias-depois-da-covid-19-provavelmente-sim,c2a68cd1be2a2c1d263e1a81cceb35edp99bik8i.html>. Acesso em: 02 de mar. 2023.

BATAGIN-NETO, A. et al. Modelling polymers with side chains: MEH-PPV and P3HT. **Molecular Simulation**, v. 39, n. 4, p. 309–321, 2013a.

BATAGIN-NETO, A.; LAVARDA, F. C. The correlation between electronic structure and antimalarial activity of alkoxyated and hydroxylated chalcones. **Medicinal Chemistry Research**, v. 23, n. 2, p. 580–586, 2013b.

BEIGEL, J. H. et al. Remdesivir for the Treatment of Covid-19 — Final Report. **New England Journal of Medicine**, v. 383, n. 19, 2020.

BREIMAN, L. Random Forests. **Machine Learning**, v. 45, n. 1, p. 5–32, 2001.

BROWN, S. D. Chemometrics: A textbook. **Journal of Chemometrics**, v. 2, n. 4, p. 298–299, 1988.

CANTÓN, R. et al. New variants of SARS-CoV 2. **Revista Española de Quimioterapia**, v. 34, n. 5, p. 419, 2021.

CAO, B. et al. A Trial of Lopinavir–Ritonavir in Adults Hospitalized with Severe Covid-19. **New England Journal of Medicine**, v. 382, n. 19, p. 1787–1799, 2020.

CENTERS FOR DISEASE CONTROL AND PREVENTION (CDC). **SARS-CoV-2 Variant Classifications and Definitions**. 2022. Disponível em: <https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-classifications.html>. Acesso em: 03/08/2022

CHAN, J. F. W. et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. **The Lancet**, v. 395, n. 10223, p. 514–523, 2020.

CHEMSKETCH, Advanced Chemistry Development Inc, Toronto, Canada. Available at: <http://www.acdlabs.com/> Acesso em: 16/03/2021

CHIRLIAN, L. E.; FRANCL, M. M. Atomic charges derived from electrostatic potentials: A detailed study. **Journal of Computational Chemistry**, v. 8, n. 6, p. 894–905, 1987.

CHUCK, C. P. et al. Profiling of Substrate Specificities of 3C-Like Proteases from Group 1, 2a, 2b, and 3 Coronaviruses. **PLoS ONE**, v. 6, n. 11, p. e27228, 2011.

CORNELL, W. D. et al. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. **Journal of the American Chemical Society**, v. 117, n. 19, p. 5179–5197, 1995.

COSAR, Begum et al. SARS-CoV-2 mutations and their viral variants. **Cytokine & growth factor reviews**, v. 63, p. 10-22, 2021.

COSSI, M. et al. Ab initio study of solvated molecules: a new implementation of the polarizable continuum model. **Chemical Physics Letters**, v. 255, n. 4-6, p. 327–335, 1996.

DAI, W. et al. Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease. **Science**, v. 368, n. 6497, p. 1331–1335, 2020.

DE PROFT, F. et al. Atomic charges, dipole moments, and Fukui functions using the Hirshfeld partitioning of the electron density. **Journal of Computational Chemistry**, v. 23, n. 12, p. 1198–1209, 2002.

DINIZ, M. O.; FERREIRA, L. C. S. Biotecnologia aplicada ao desenvolvimento de vacinas. **Estudos Avançados**, v. 24, n. 70, p. 2010, 2010.

FERRAZ, L. G. W. **MINISTÉRIO DA ECONOMIA INSTITUTO NACIONAL DA PROPRIEDADE INDUSTRIAL VACINAS À BASE DE SUBUNIDADE PROTEICA PARA PREVENÇÃO DA COVID-19: Mecanismo de ação, ensaios clínicos e pedidos de patentes**. 2021 Disponível em: <https://www.gov.br/inpi/pt-br/servicos/patentes/tecnologias-para-covid-19/Arquivos%20Textos/estudo10-de-vacinas-suunidade_03052021.pdf>. Acesso em: 03/08/2022.

FERREIRA, M. M. C. Multivariate QSAR. **Journal of the Brazilian Chemical Society**, v. 13, n. 6, p. 742–753, 2002.

FIOROTTO DE MATTOS, A. Cloroquina e Hidroxicloroquina: Seus Efeitos no Tratamento da COVID-19. **Ensaios e Ciência: Ciências Biológicas Agrárias e da Saúde**, v. 25, n. 4, p. 468–472, 2021.

FRISCH, M. J. et al. **Gaussian 16 Revision A.03**. Gaussian Inc. Wallingford CT, 2016.

GENUER, R.; POGGI, J.-M.; TULEAU-MALOT, C. Variable selection using random forests. *Pattern Recognition Letters*, v. 31, n. 14, p. 2225–2236, out. 2010.

GROSSE, M. et al. Quinine Inhibits Infection of Human Cell Lines with SARS-CoV-2. **Viruses**, v. 13, n. 4, p. 647, 2021.

HALGREN, T. Performance of MMFF94*. **Scope, Parameterization, and Journal of Computational Chemistry**, v. 17, p. 490–519, 1996.

HANWELL, M. D. et al. Avogadro: an advanced semantic chemical editor,

visualization, and analysis platform. **Journal of Cheminformatics**, v. 4, n. 1, 13 ago. 2012.

HARTENIAN, E. et al. The molecular virology of coronaviruses. **Journal of Biological Chemistry**, v. 295, n. 37, p. 12910–12934, 2020.

INSTITUTO BUTANTAN. **Número de mortes globais por Covid-19 pode ser três vezes maior do que os registros oficiais, indica estudo americano**. 2022.

Disponível em:

<https://butantan.gov.br/noticias/numero-de-mortes-globais-por-covid-19-pode-ser-tres-vezes-maior-do-que-os-registros-oficiais-indica-estudo-americano>. Acesso em: 26/07/2022

INSTITUTO BUTANTAN. **Depois da Covid-19, será que estamos preparados para novas pandemias?** 2021. Disponível em:

<https://butantan.gov.br/noticias/depois-da-covid-19-sera-que-estamos-preparados-para-novas-pandemias->. Acesso em: 28 fev. 2023.

KRAMMER, F. SARS-CoV-2 vaccines in development. **Nature**, v. 586, n. 7830, p. 516–527, 2020.

LAMERS, M. M.; HAAGMANS, B. L. SARS-CoV-2 pathogenesis. **Nature Reviews Microbiology**, v. 20, p. 1–15, 30 mar. 2022.

LI, J. et al. Structural Basis of the Main Proteases of Coronavirus Bound to Drug Candidate PF-07321332. **Journal of Virology**, v. 96, n. 8, p. e0201321, 27 abr. 2022.

LIMA, E. J. F.; ALMEIDA, A. M.; KFOURI, R. Á. Vacinas para COVID-19: o estado da arte. **Revista Brasileira de Saúde Materno Infantil**, v. 21, p. 13–19, 2021.

MANIR, M. **Coronavírus pode ser só “ensaio” de uma próxima grande pandemia, diz médico e matemático da USP**. BBC News Brasil. 2020. Disponível em <https://www.bbc.com/portuguese/geral-52389645> Acesso em: 02/03/2023.

MARDIROSSIAN, N.; HEAD-GORDON, M. Thirty years of density functional theory in computational chemistry: an overview and extensive assessment of 200 density functionals. **Molecular Physics**, v. 115, n. 19, p. 2315–2372, 2017.

MARTÍNEZ, A. et al. Donator Acceptor Map for Carotenoids, Melatonin and Vitamins. **The Journal of Physical Chemistry A**, v. 112, n. 38, p. 9037–9042, 2008.

NARANJO-MONTOYA, O. A. et al. The Correlation Between Electronic Structure and Antimalarial Activity of Tetrahydropyridines. **Journal of the Brazilian Chemical Society**, v. 26, n. 2, p. 255–265, 2014.

NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION. **PubChem Compound Summary for CID 155903259**, Nirmatrelvir. 2023. Disponível em: <<https://pubchem.ncbi.nlm.nih.gov/compound/Nirmatrelvir>>. Acesso em: 09 mar. 2023.

NAVECA, F. G. et al. COVID-19 in Amazonas, Brazil, was driven by the persistence of endemic lineages and P.1 emergence. **Nature Medicine**, v. 27, n. 7, p. 1230–1238, 2021.

NICOLETI, N. H.; BATAGIN-NETO, A.; LAVARDA, F. C. Electronic descriptors for the antimalarial activity of sulfonamides. **Medicinal Chemistry Research**, v. 25, n. 8, p. 1630–1638, 2016.

PAXLOVID: Comprimido. [Bula] Responsável técnico Andrea T. Nichele. São Paulo: Pfizer Brasil LTDA., 2022. Disponível em: https://www.pfizer.com.br/files/Paxlovid_Profissional_de_Saude_12.pdf. Acesso em: 7 mar. 2023.

PEDREGOSA, F. et al. Scikit-learn: Machine Learning in Python Gaël Varoquaux Bertrand Thirion Vincent Dubourg Alexandre Passos PEDREGOSA, VAROQUAUX,

GRAMFORT ET AL. Matthieu Perrot Edouard Duchesnay. **Journal of Machine Learning Research**, v. 12, p. 2825–2830, 2011.

PIZZORNO, A. et al. In vitro evaluation of antiviral activity of single and combined repurposable drugs against SARS-CoV-2. **Antiviral Research**, v. 181, p. 104878, 2020.

RAMOS-GUZMÁN, C. A.; RUIZ-PERNÍA, J. J.; TUÑÓN, I. Multiscale Simulations of SARS-CoV-2 3CL Protease Inhibition with Aldehyde Derivatives. Role of Protein and Inhibitor Conformational Changes in the Reaction Mechanism. **ACS Catalysis**, v. 11, n. 7, p. 4157–4168, 25 mar. 2021.

RODRIGUES, Ricardo P. et al. Estratégias de triagem virtual no planejamento de fármacos. *Revista Virtual de Química*, v. 4, n. 6, p. 739-776, 2012.

ROY, R. K.; PAL, S.; HIRAO, K. On non-negativity of Fukui function indices. *The Journal of Chemical Physics*, v. 110, n. 17, p. 8236–8245, 1999.

SCHAFTENAAR, G.; NOORDIK, J. H. Molden: a pre- and post-processing program for molecular and electronic structures. **Journal of Computer-Aided Molecular Design**, v. 14, n. 2, p. 123–134, 2000.

SILVA, J. L. et al. Introdução a Triagem Virtual. **BIOINFO - Revista Brasileira de Bioinformática e Biologia Computacional**, jul. 2021.

STEWART, J. J. P. Application of the PM6 method to modeling the solid state. **Journal of Molecular Modeling**, v. 14, n. 6, p. 499–535, 2008.

STEWART, J. P. **Mopac2016**. Stewart Computational Chemistry: Colorado Springs, CO, USA, v. 650, 2016.

TEAM, T. A. **Molecular Mechanics & Force Fields**. Disponível em:

<<https://avogadro.cc/docs/optimizing-geometry/molecular-mechanics/>>. Acesso em: 27/07/2022.

VENABLES, W. N.; RIPLEY, B. D. **Modern Applied Statistics with S**. New York, NY: Springer New York, 2002.

WANG, H. et al. Estimating excess mortality due to the COVID-19 pandemic: a systematic analysis of COVID-19-related mortality, 2020–21. **The Lancet**, v. 399, n. 10334, 2022.

WANG, J. et al. Development and testing of a general amber force field. **Journal of Computational Chemistry**, v. 25, n. 9, p. 1157–1174, 2004.

WANG, M. et al. Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro. **Cell Research**, v. 30, n. 3, p. 269–271, 2020.

WORLD ECONOMIC SITUATION AND PROSPECTS (WESP). **Situação Econômica Mundial e Perspectivas**: Briefing de fevereiro de 2021, No. 146.

Disponível em:

<https://www.un.org/development/desa/dpad/publication/world-economic-situation-and-prospects-february-2021-briefing-no-146/> Acesso em: 03/08/2021

XAVIER, A. R. et al. COVID-19: clinical and laboratory manifestations in novel coronavirus infection. **Jornal Brasileiro de Patologia e Medicina Laboratorial**, v.34, p.101623, 2020.

YANG, H. et al. Design of Wide-Spectrum Inhibitors Targeting Coronavirus Main Proteases. **PLoS Biology**, v. 3, n. 10, p. e324, 2005.

YANG, W.; MORTIER, W. J. The use of global and local molecular parameters for the analysis of the gas-phase basicity of amines. **Journal of the American Chemical Society**, v. 108, n. 19, p. 5708–5711, 1986.

ZHANG, L. et al. α -Ketoamides as broad-spectrum inhibitors of coronavirus and enterovirus replication: structure-based design, synthesis, and activity assessment. **Journal of medicinal chemistry**, v. 63, n. 9, p. 4562-4578, 2020a.

ZHANG, L. et al. Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α -ketoamide inhibitors. **Science**, v. 368, n. 6489, p. 409-412, 2020b.

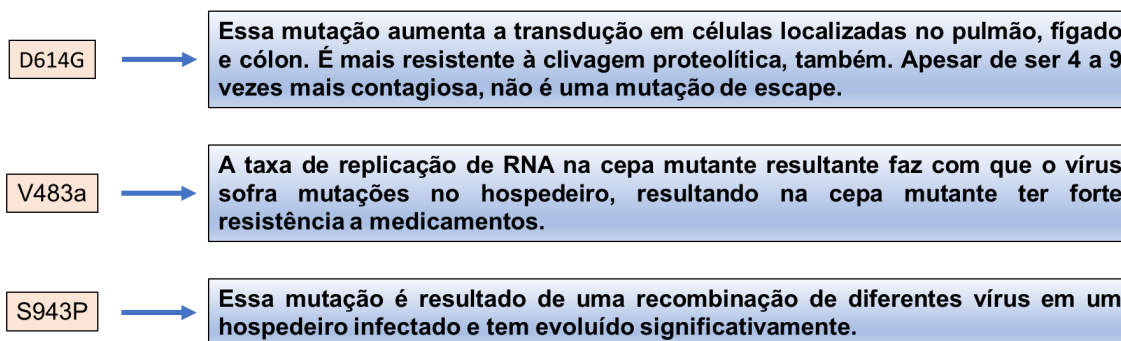
ZHOU, P. et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. **Nature**, v. 579, n. 7798, p. 270–273, 2020.

ZHU, N. et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. **New England Journal of Medicine**, v. 382, n. 8, p. 727–733, 2020.

ANEXO A

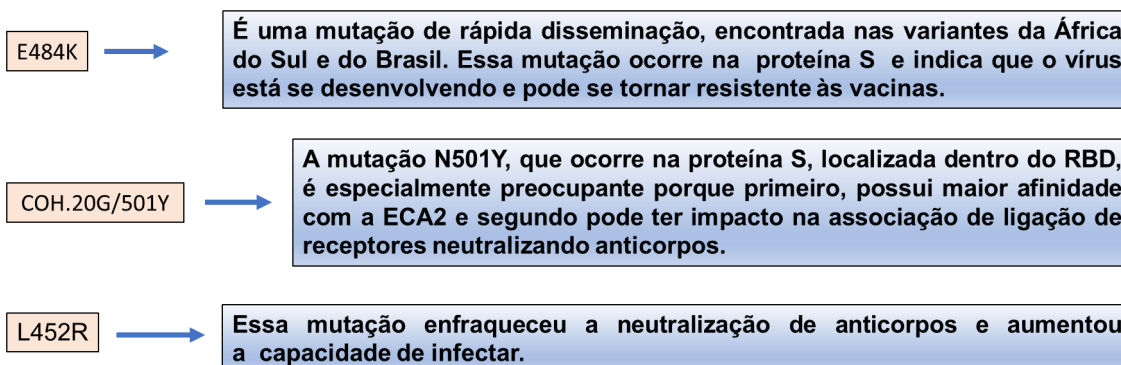
As Figuras A1-A3 representam uma breve explicação sobre as principais mutações sofridas pelo SARS-CoV-2.

Figura A1. Mutações do SARS-CoV-2: D614G, V483a, S943P.



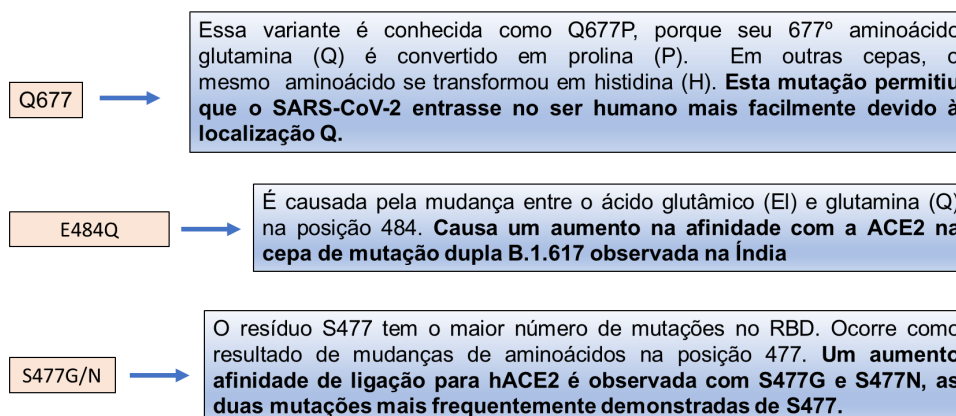
Fonte: Adaptado de Cosar et al. (2021)

Figura A2. Mutações do SARS-CoV-2: E484K, N501Y, L452.



Fonte: Adaptado de Cosar et al. (2021)

Figura A3. Mutações do SARS-CoV-2: Q677, E484Q, S477G/N.

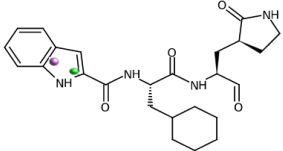
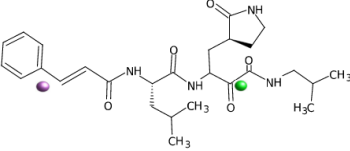
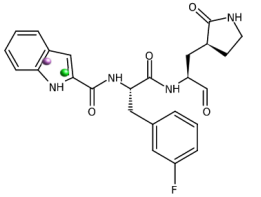
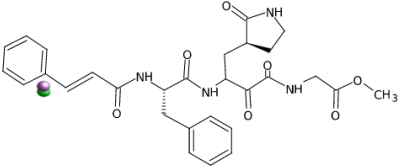
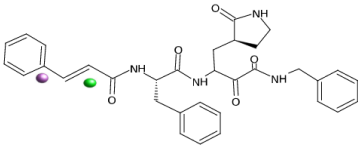
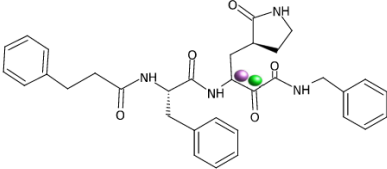
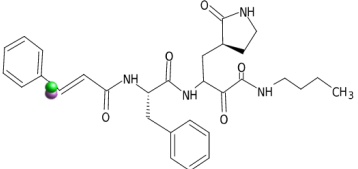
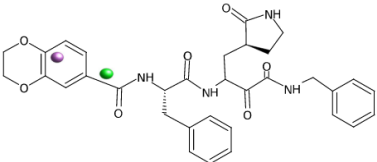
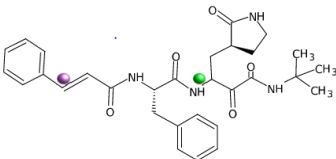
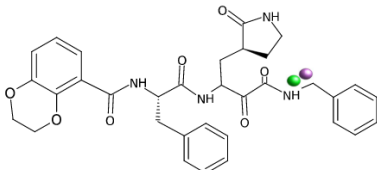


Fonte: Adaptado de Cosar et al. (2021)

APÊNDICE I

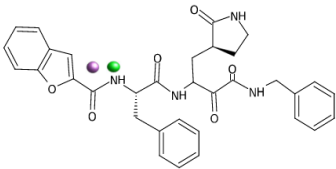
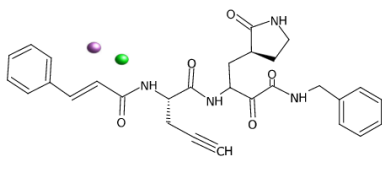
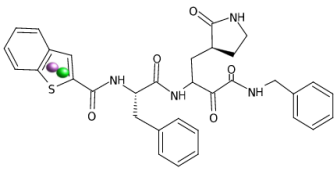
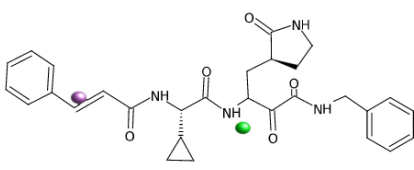
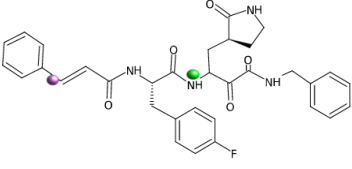
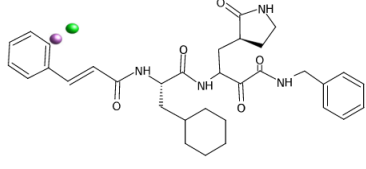
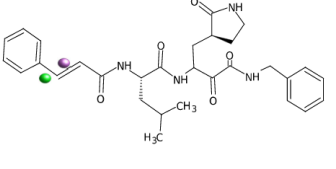
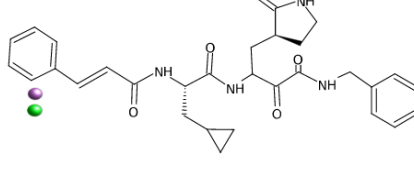
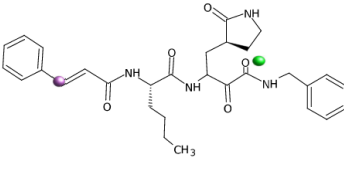
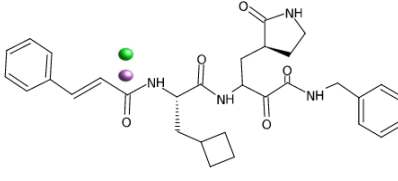
As Tabelas A1-A3 apresentam representações em duas dimensões das posições dos centróides associados à f^+ (verde) e f^- (magenta) em relação às estruturas dos derivados presentes na literatura. Tais representações foram obtidas com o auxílio do software SAMSON e LibreOffice Draw.

Tabela A1. Representação bidimensional da posição dos centróides associados à f^+ (verde) e f^- (magenta) em relação a estrutura dos compostos: 11a1, 11a2, 11b1, 11c, 11d, 11e, 11g, 11h, 11i.

Composto	Centroids 2D	Composto	Centroids 2D
11a1		11d	
11b1		11e	
11a2		11g	
11b2		11h	
11c		11i	

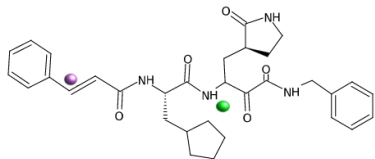
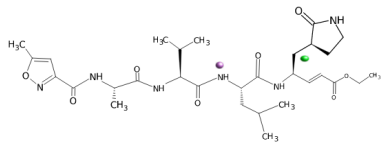
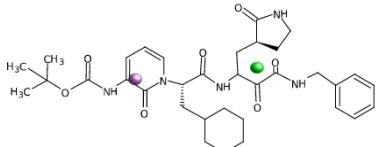
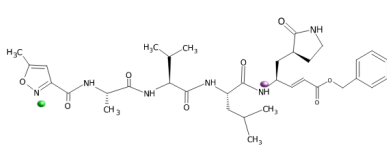
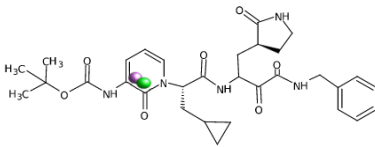
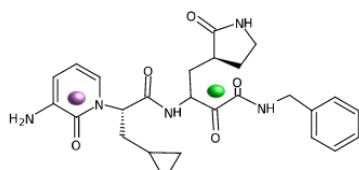
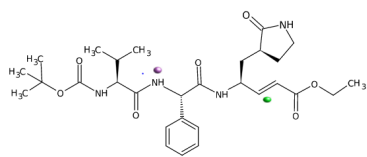
Fonte: Autoria própria.

Tabela A2. Representação bidimensional da posição dos centróides associados à f⁺ (verde) e f⁻ (magenta) em relação a estrutura dos compostos: 11j, 11k, 11m, 11n, 11o, 11p, 11q, 11r, 11s, 11t

Composto	Centroids 2D	Composto	Centroids 2D
11j		11p	
11k		11q	
11m		11r	
11n		11s	
11o		11t	

Fonte: Autoria própria.

Tabela A3. Representação bidimensional da posição dos centróides associados à f^+ (verde) e f^- (magenta) em relação a estrutura dos compostos: 11u, 13a, 13b, 14b, I2, N1, N3.

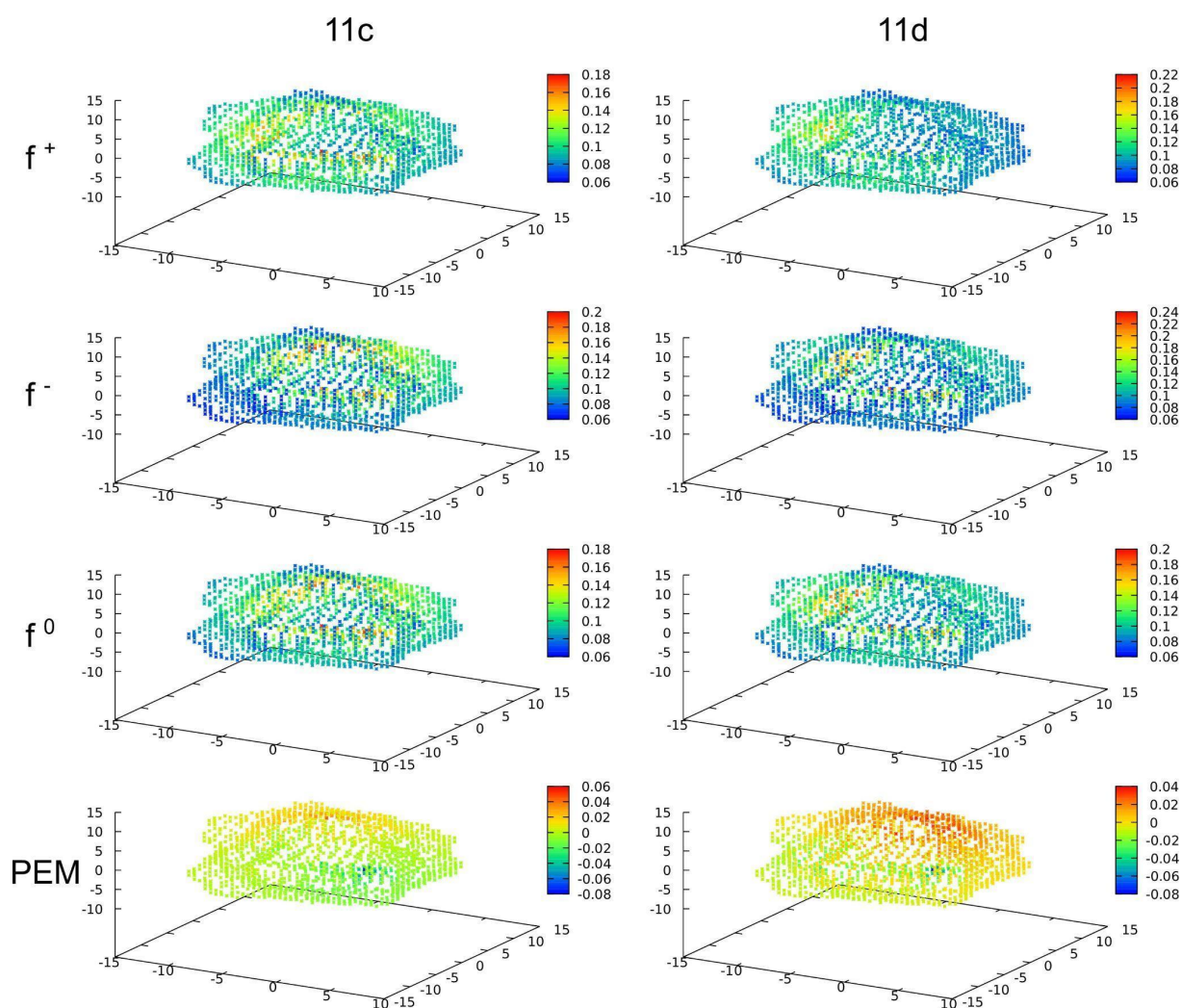
Composto	Centroids 2D	Composto	Centroids 2D
11u		N1	
13a		N3	
13b			
14b			
I2			

Fonte: Autoria própria.

APÊNDICE II

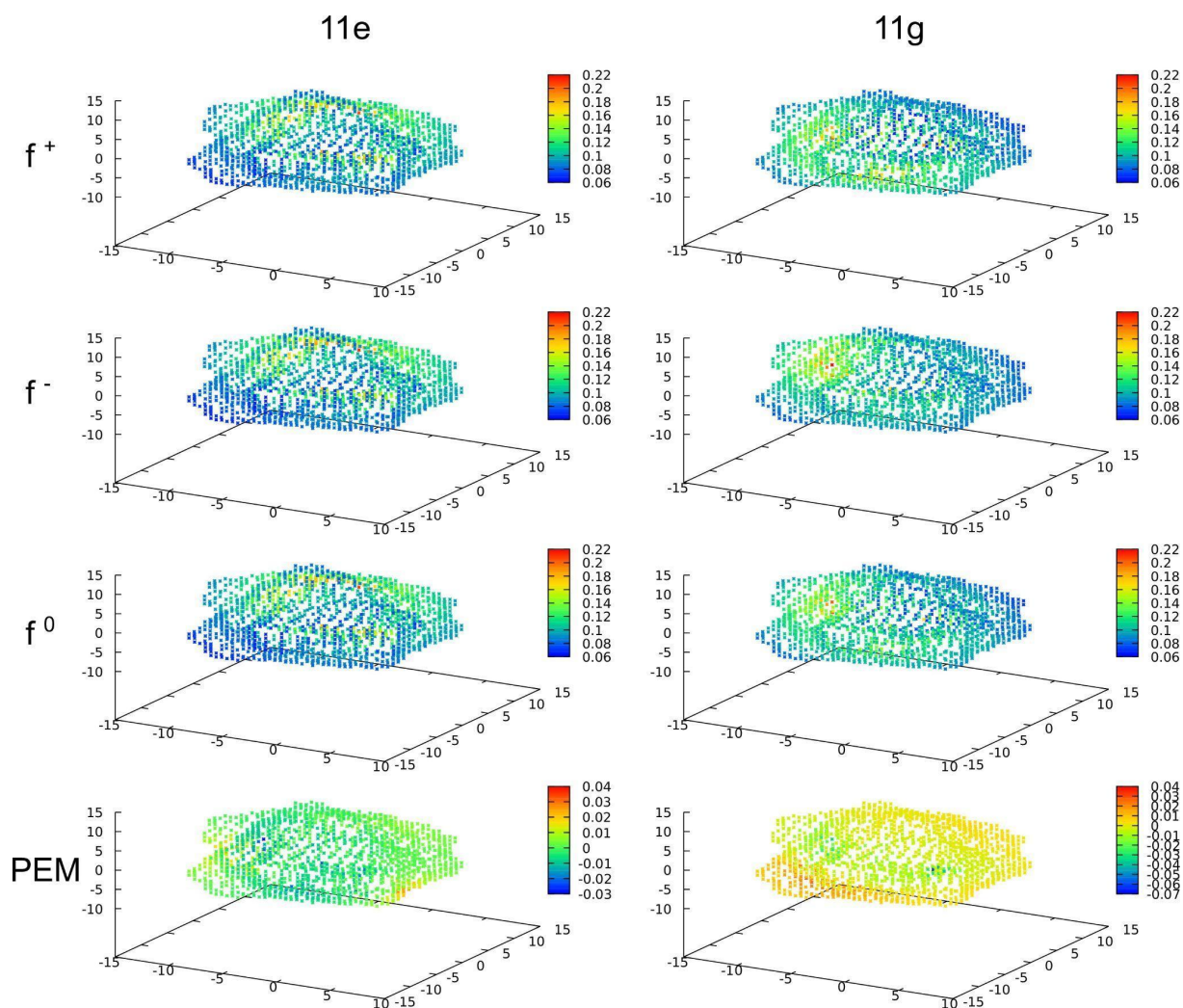
As Figuras A4-A14 representam as superfícies moleculares 3D dos compostos 11c, 11d, 11e, 11g, 11h, 11i, 11j, 11k, 11o, 11p, 11q, 11r, 11s, 11t, 11u, 13a, 13b, 14b, I2, N1 e N3, dois a dois.

Figura A4. Representação das Superfícies Moleculares 3D com os descritores IFCA e PEM: 11c e 11d. Os eixos coordenados x, y e z, representam as coordenadas espaciais em Å (omitidas por simplicidade).



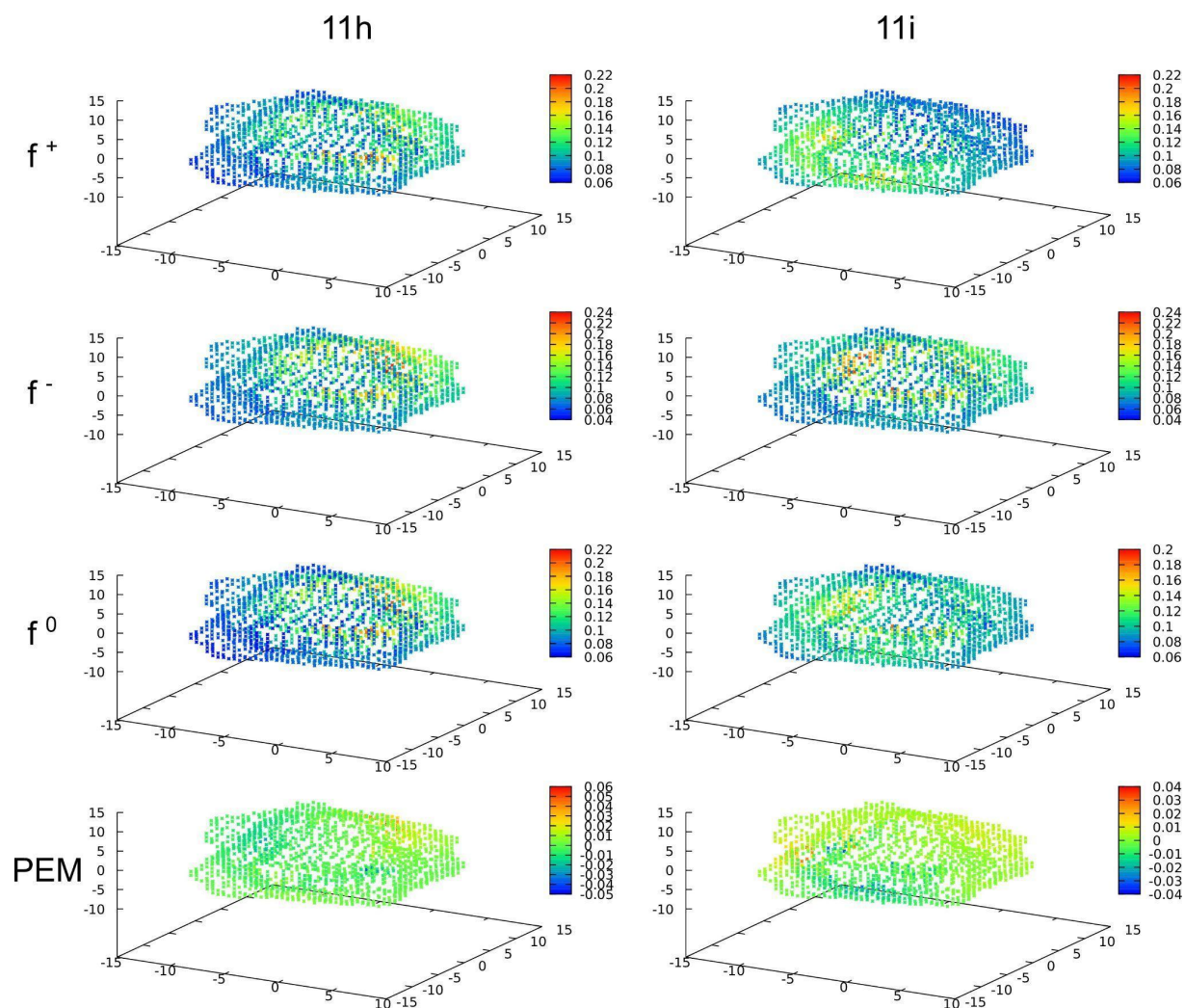
Fonte: Autoria própria.

Figura A5. Representação das Superfícies Moleculares 3D com os descritores IFCA e PEM: 11e e 11g. Os eixos coordenados x, y e z, representam as coordenadas espaciais em Å (omitidas por simplicidade).



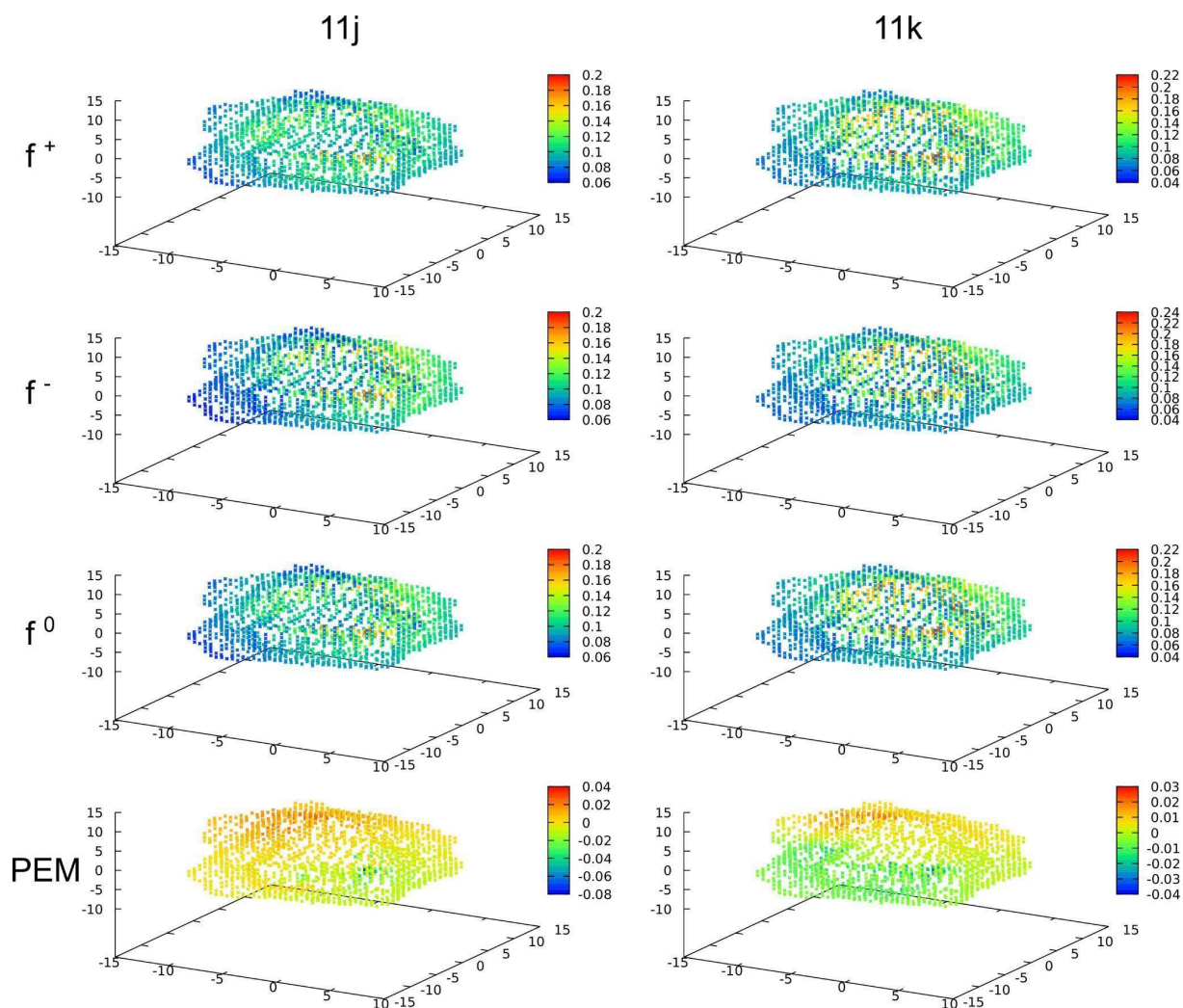
Fonte: Autoria própria.

Figura A6. Representação das Superfícies Moleculares 3D com os descritores IFCA e PEM: 11h e 11i. Os eixos coordenados x, y e z, representam as coordenadas espaciais em Å (omitidas por simplicidade).



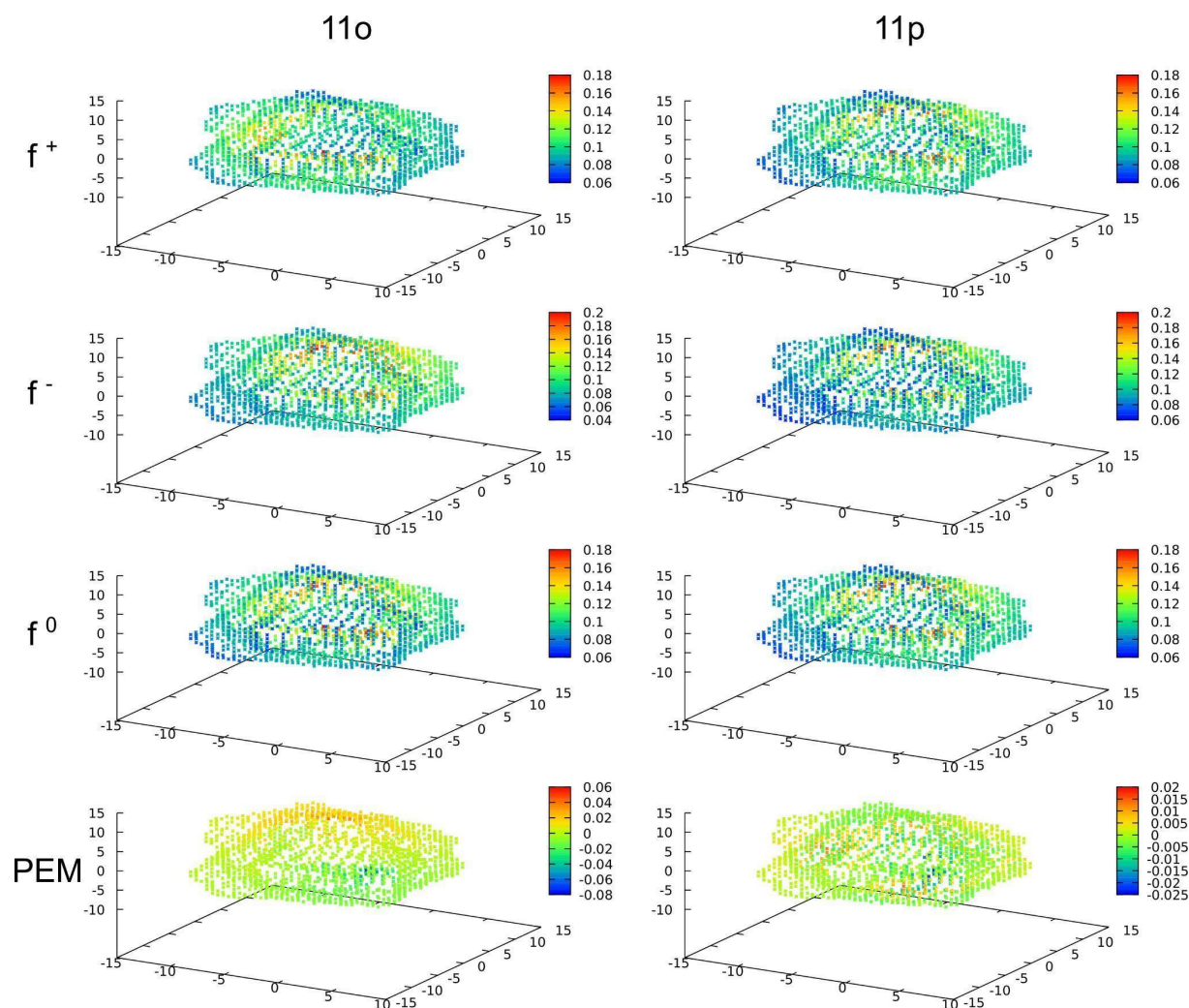
Fonte: Autoria própria.

Figura A7. Representação das Superfícies Moleculares 3D com os descritores IFCA e PEM: 11j e 11k. Os eixos coordenados x, y e z, representam as coordenadas espaciais em Å (omitidas por simplicidade).



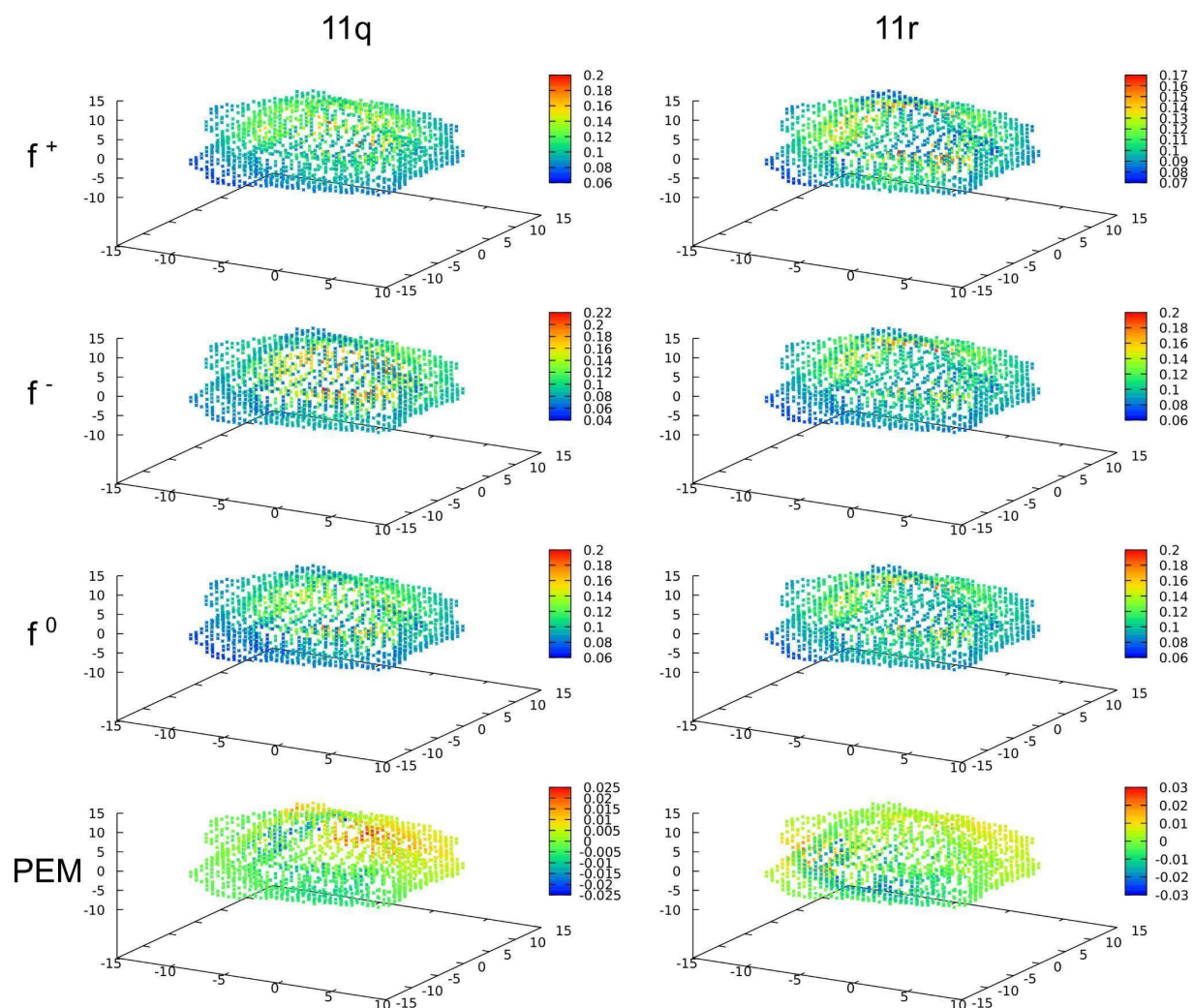
Fonte: Autoria própria.

Figura A8. Representação das Superfícies Moleculares 3D com os descritores IFCA e PEM: 11o e 11p. Os eixos coordenados x, y e z, representam as coordenadas espaciais em Å (omitidas por simplicidade).



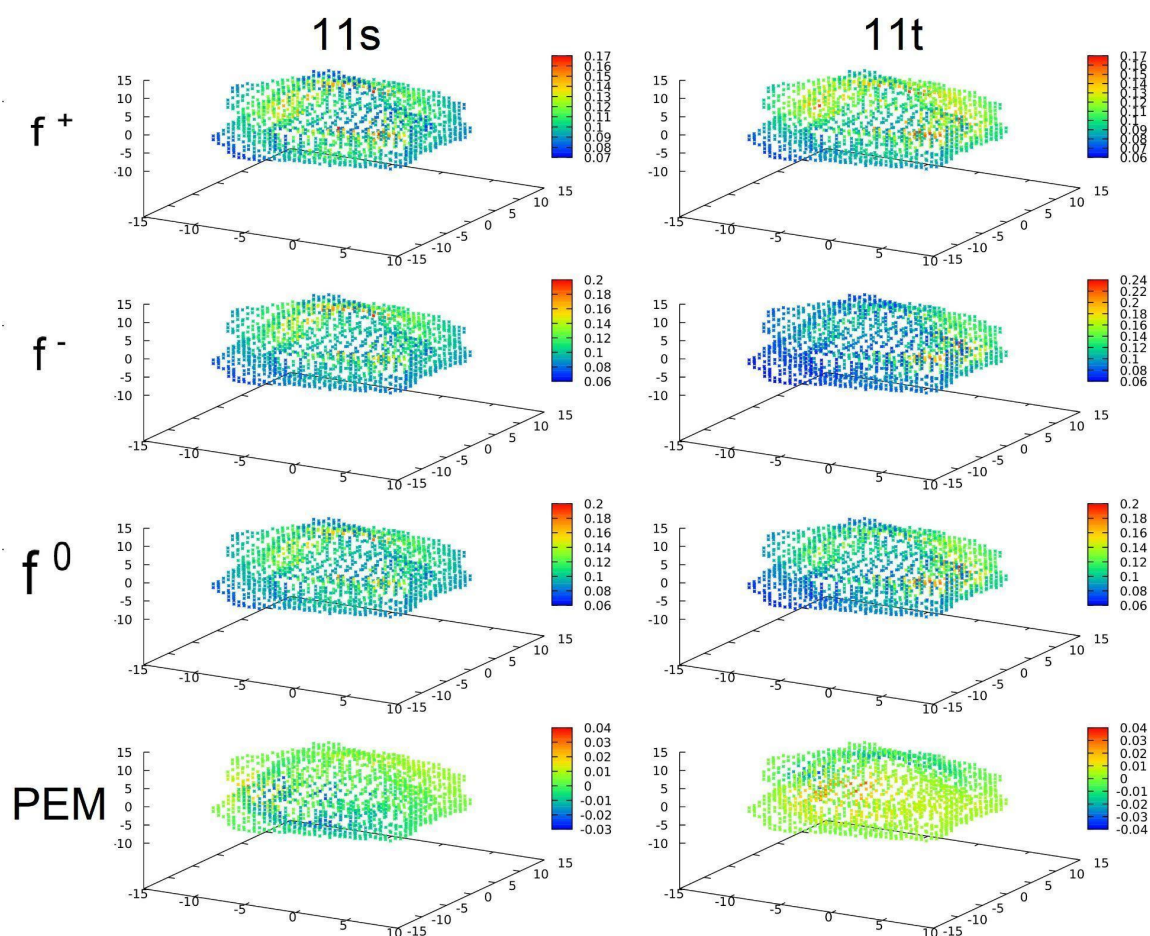
Fonte: Autoria própria.

Figura A9. Representação das Superfícies Moleculares 3D com os descritores IFCA e PEM: 11q e 11r. Os eixos coordenados x, y e z, representam as coordenadas espaciais em Å (omitidas por simplicidade).



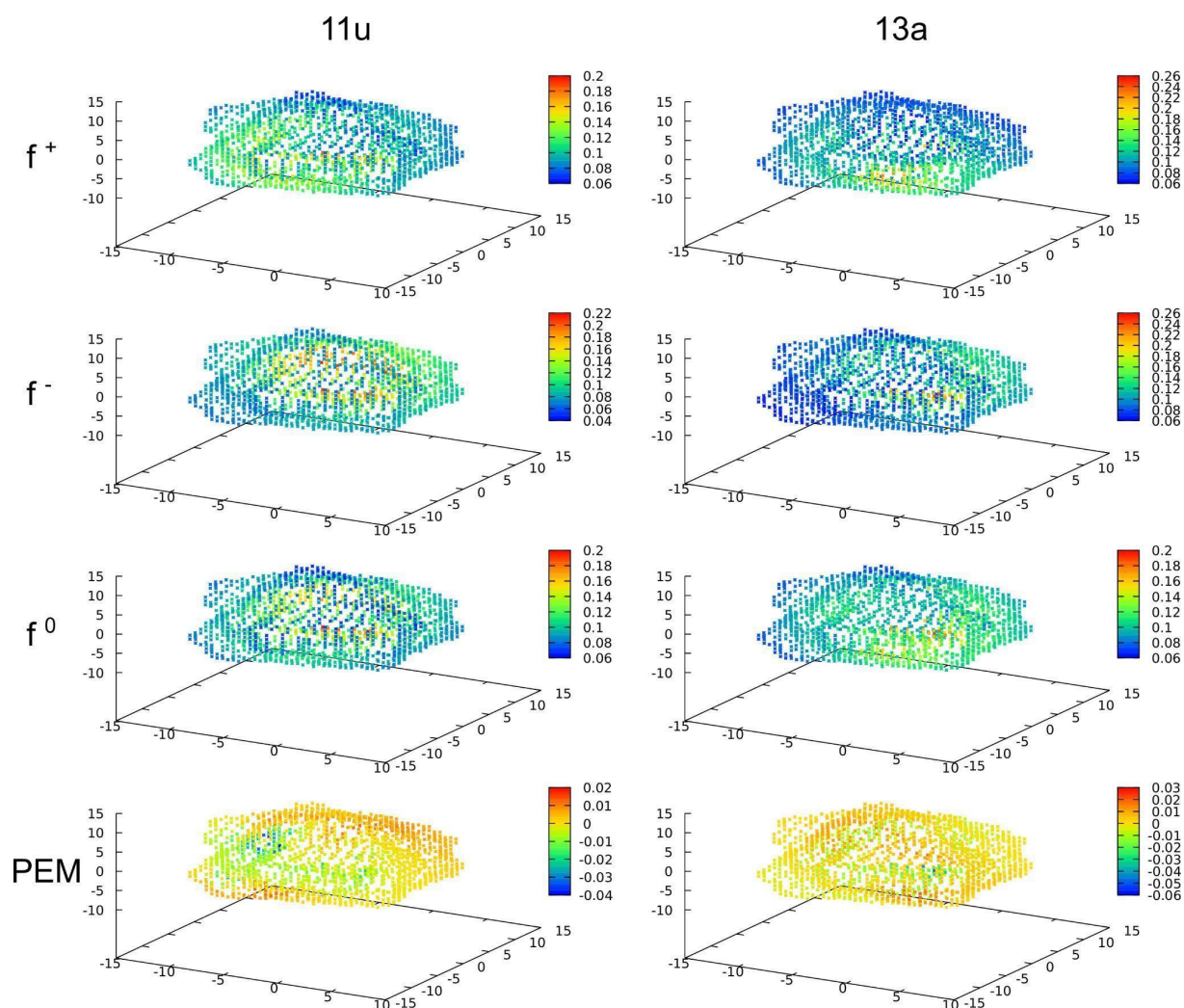
Fonte: Autoria própria.

Figura A10. Representação das Superfícies Moleculares 3D com os descritores IFCA e PEM: 11s e 11t. Os eixos coordenados x, y e z, representam as coordenadas espaciais em Å (omitidas por simplicidade).



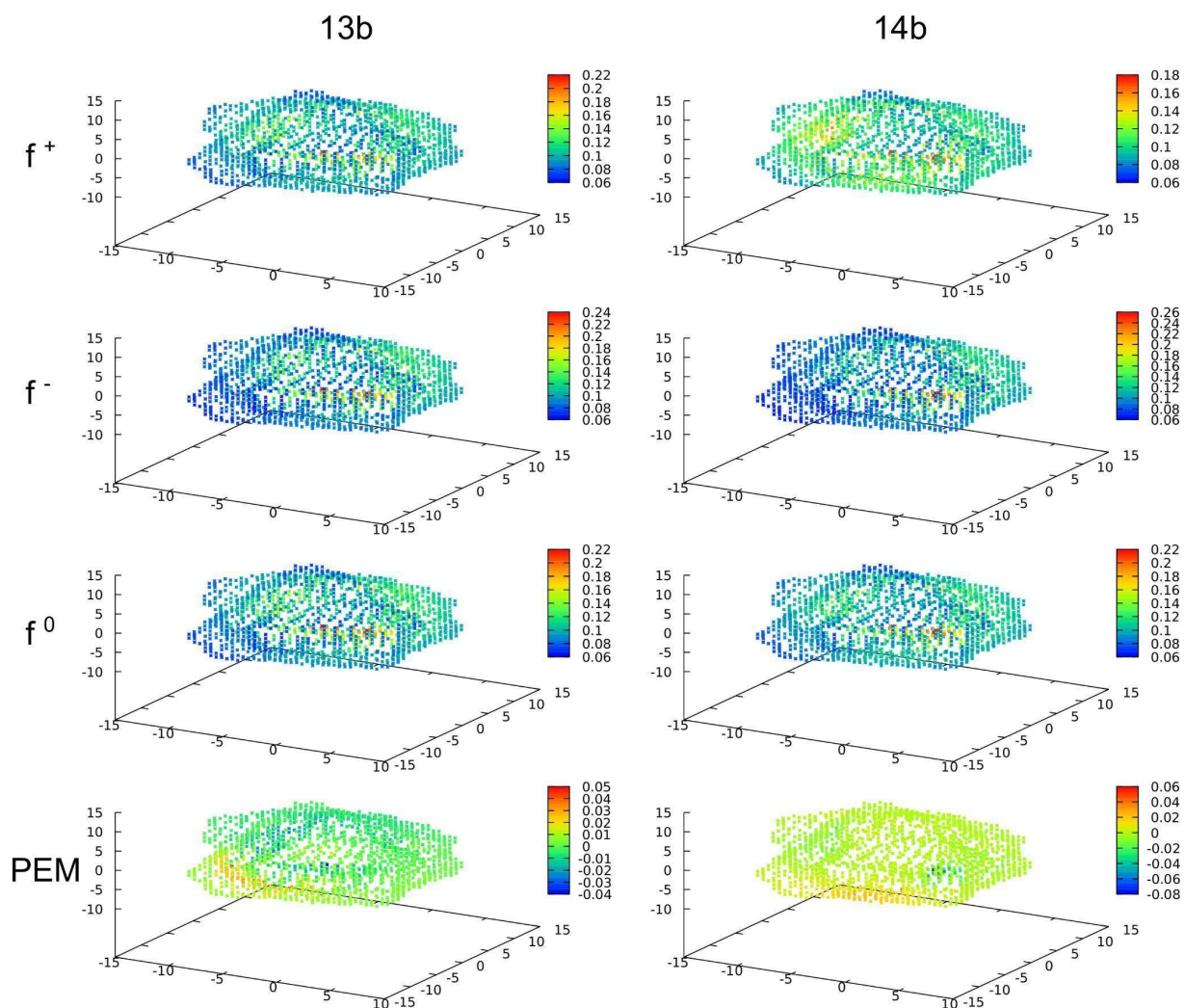
Fonte: Autoria própria.

Figura A11. Representação das Superfícies Moleculares 3D com os descritores IFCA e PEM: 11u e 13a. Os eixos coordenados x, y e z, representam as coordenadas espaciais em Å (omitidas por simplicidade).



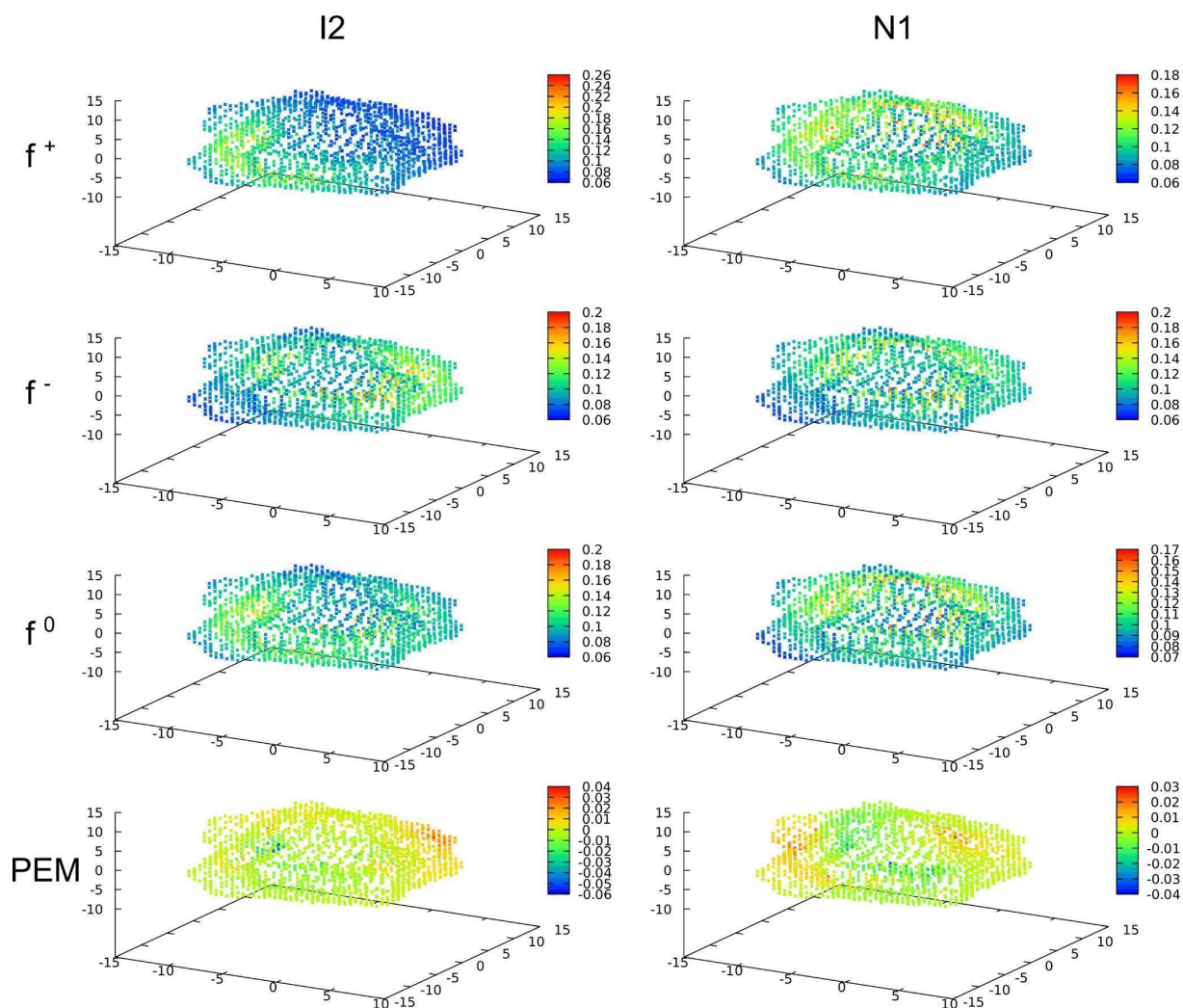
Fonte: Autoria própria.

Figura A12. Representação das Superfícies Moleculares 3D com os descritores IFCA e PEM: 13b e 14b. Os eixos coordenados x, y e z, representam as coordenadas espaciais em Å (omitidas por simplicidade).



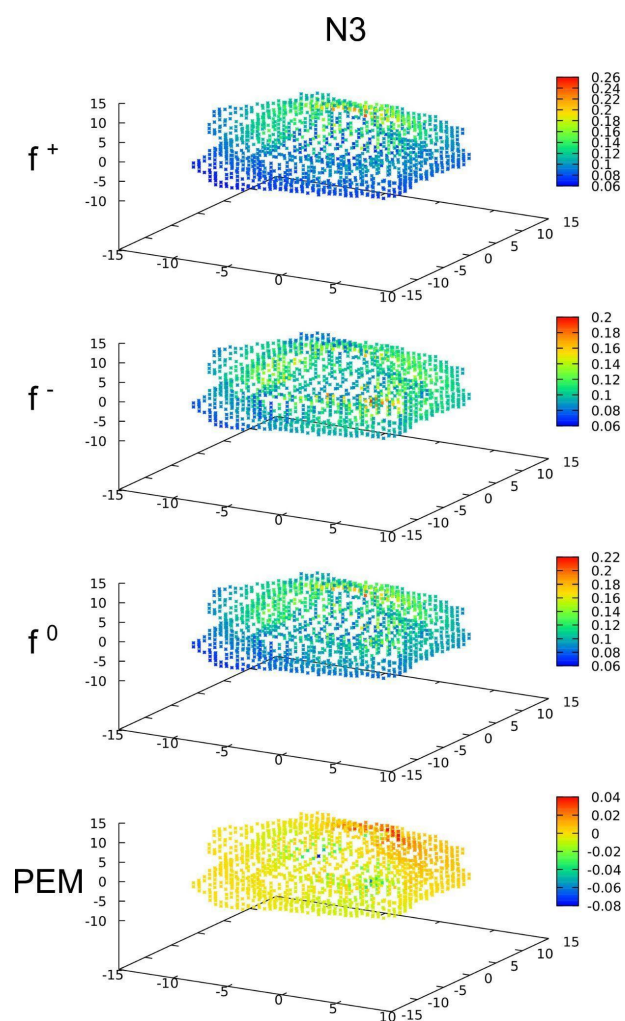
Fonte: Autoria própria.

Figura A13. Representação das Superfícies Moleculares 3D com os descritores IFCA e PEM: I2 e N1. Os eixos coordenados x, y e z, representam as coordenadas espaciais em Å (omitidas por simplicidade).



Fonte: Autoria própria.

Figura A14. Representação da Superfície Molecular 3D com os descritores IFCA e PEM: N3. Os eixos coordenados x, y e z, representam as coordenadas espaciais em Å (omitidas por simplicidade).

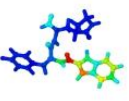
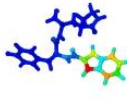
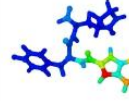
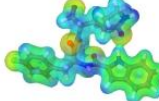
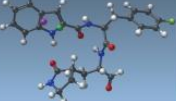
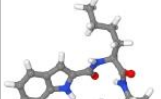
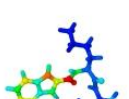
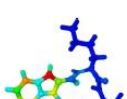
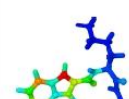
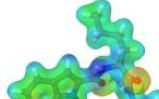
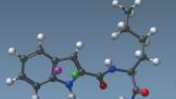
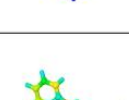
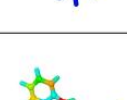
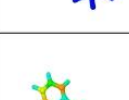
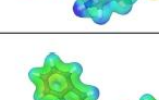
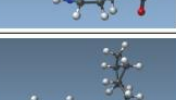
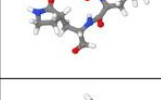
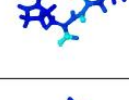
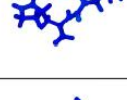
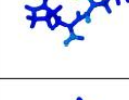
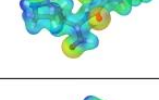
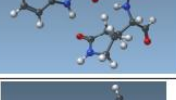
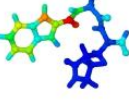
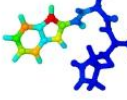
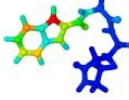
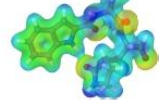
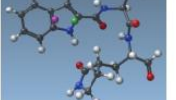
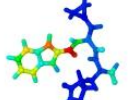
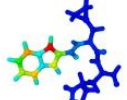
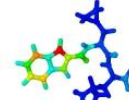
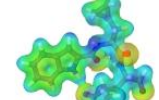
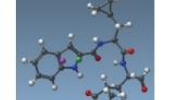
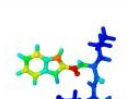
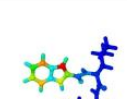
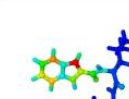
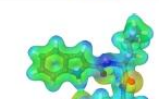
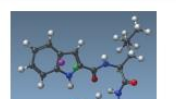


Fonte: Autoria própria.

APÊNDICE III

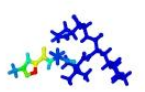
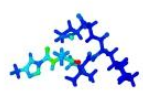
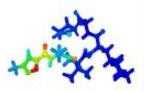
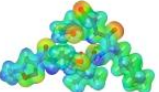
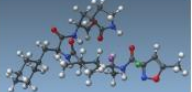
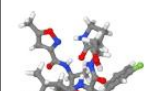
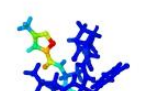
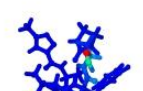
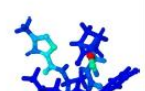
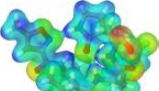
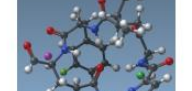
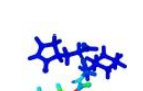
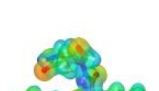
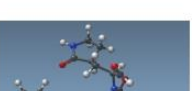
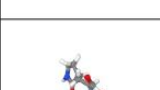
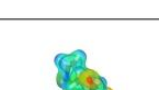
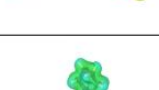
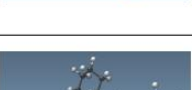
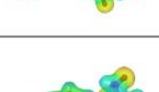
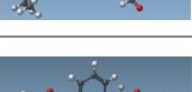
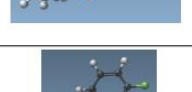
As Figuras A15-A20 representam os IFCA, PEM e centroides tridimensionais em relação aos grupos de testagem A, B, C, D e E.

Figura A15. IFCA, PEM e Centróides dos compostos do grupo de testagem: A1-A7.

Composto	Estrutura	f^+	f^-	f^0	PEM	Centroids 3D
A1						
A2						
A3						
A4						
A5						
A6						
A7						

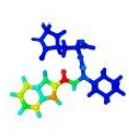
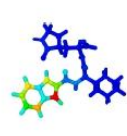
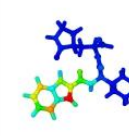
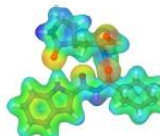
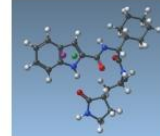
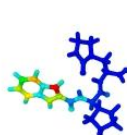
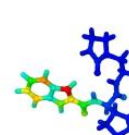
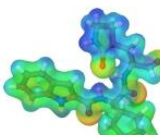
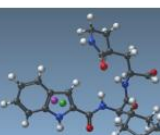
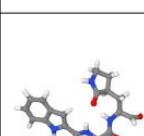
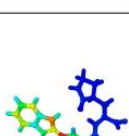
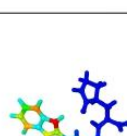
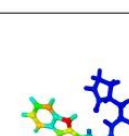
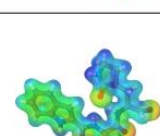
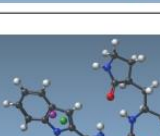
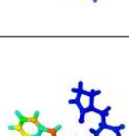
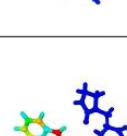
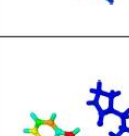
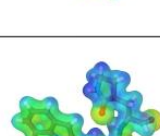
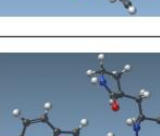
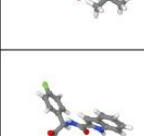
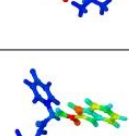
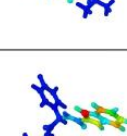
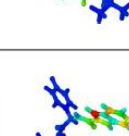
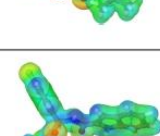
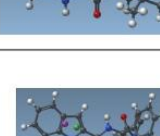
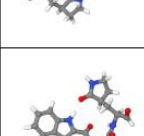
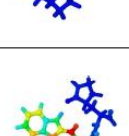
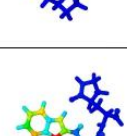
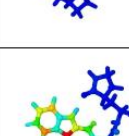
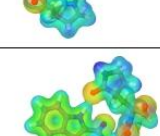
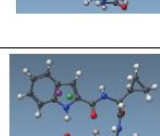
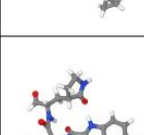
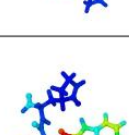
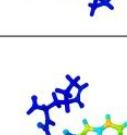
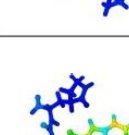
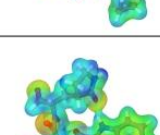
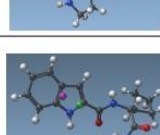
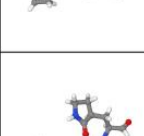
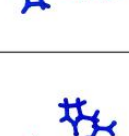
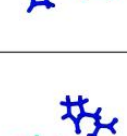
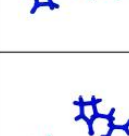
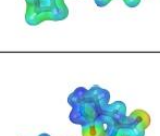
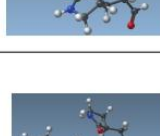
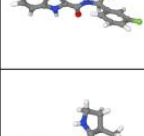
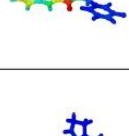
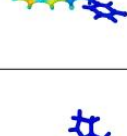
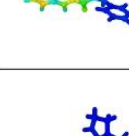
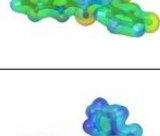
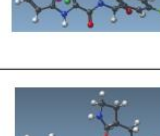
Fonte: Autoria própria.

Figura A16. IFCA, PEM e Centróides dos compostos do grupo de testagem: B1-B8.

Composto	Estrutura	f^+	f^-	f^0	PEM	Centroids 3D
B1						
B2						
B3						
B4						
B5						
B6						
B7						
B8						

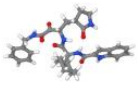
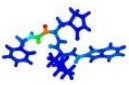
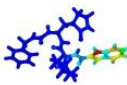
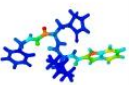
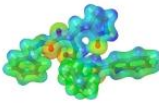
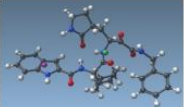
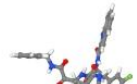
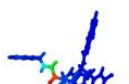
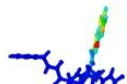
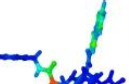
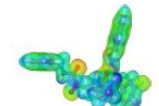
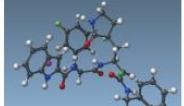
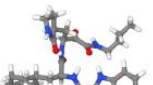
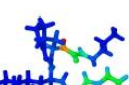
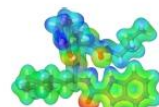
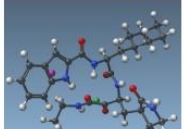
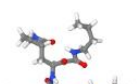
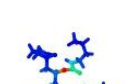
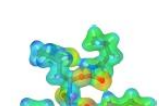
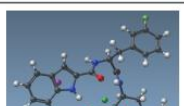
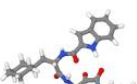
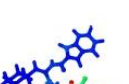
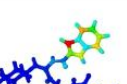
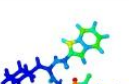
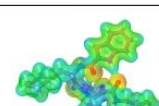
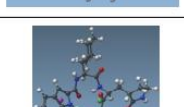
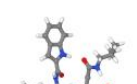
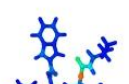
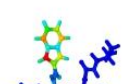
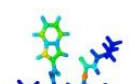
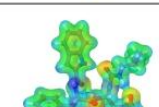
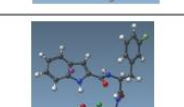
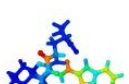
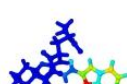
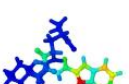
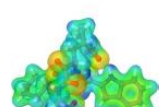
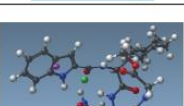
Fonte: Autoria própria.

Figura A17. IFCA, PEM e Centróides dos compostos do grupo de testagem: C1-C9.

Composto	Estrutura	f^+	f^-	f^0	PEM	Centroids 3D
C1						
C2						
C3						
C4						
C5						
C6						
C7						
C8						
C9						

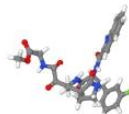
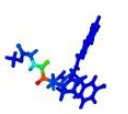
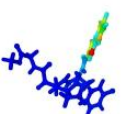
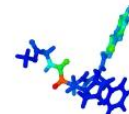
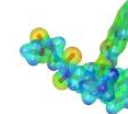
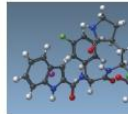
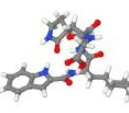
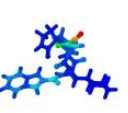
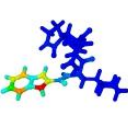
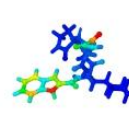
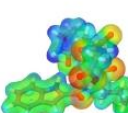
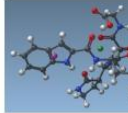
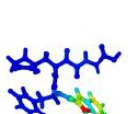
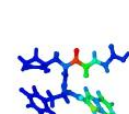
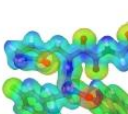
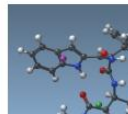
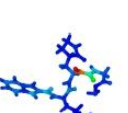
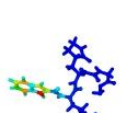
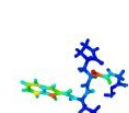
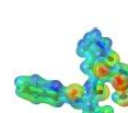
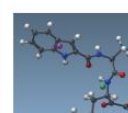
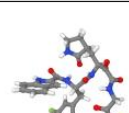
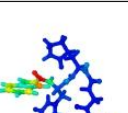
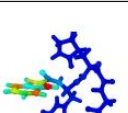
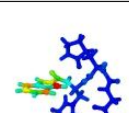
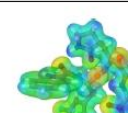
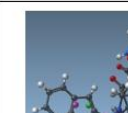
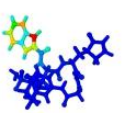
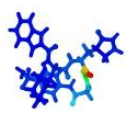
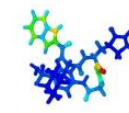
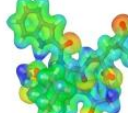
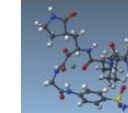
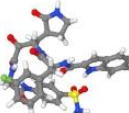
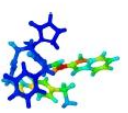
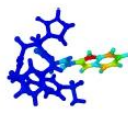
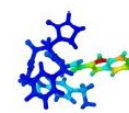
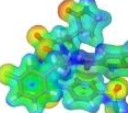
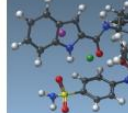
Fonte: Autoria própria.

Figura A18. IFCA, PEM e Centróides dos compostos do grupo de testagem: D1-D7.

Composto	Estrutura	f^+	f^-	f^0	PEM	Centroids 3D
D1						
D2						
D3						
D4						
D5						
D6						
D7						

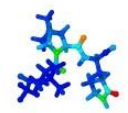
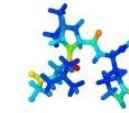
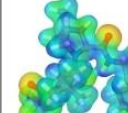
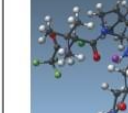
Fonte: Autoria própria.

Figura A19. IFCA, PEM e Centróides dos compostos do grupo de testagem: D8-D14.

Composto	Estrutura	f^+	f^-	f^0	PEM	Centroids 3D
D8						
D9						
D10						
D11						
D12						
D13						
D14						

Fonte: Autoria própria.

Figura A20. IFCA, PEM e Centróides do Paxlovid, do grupo E.

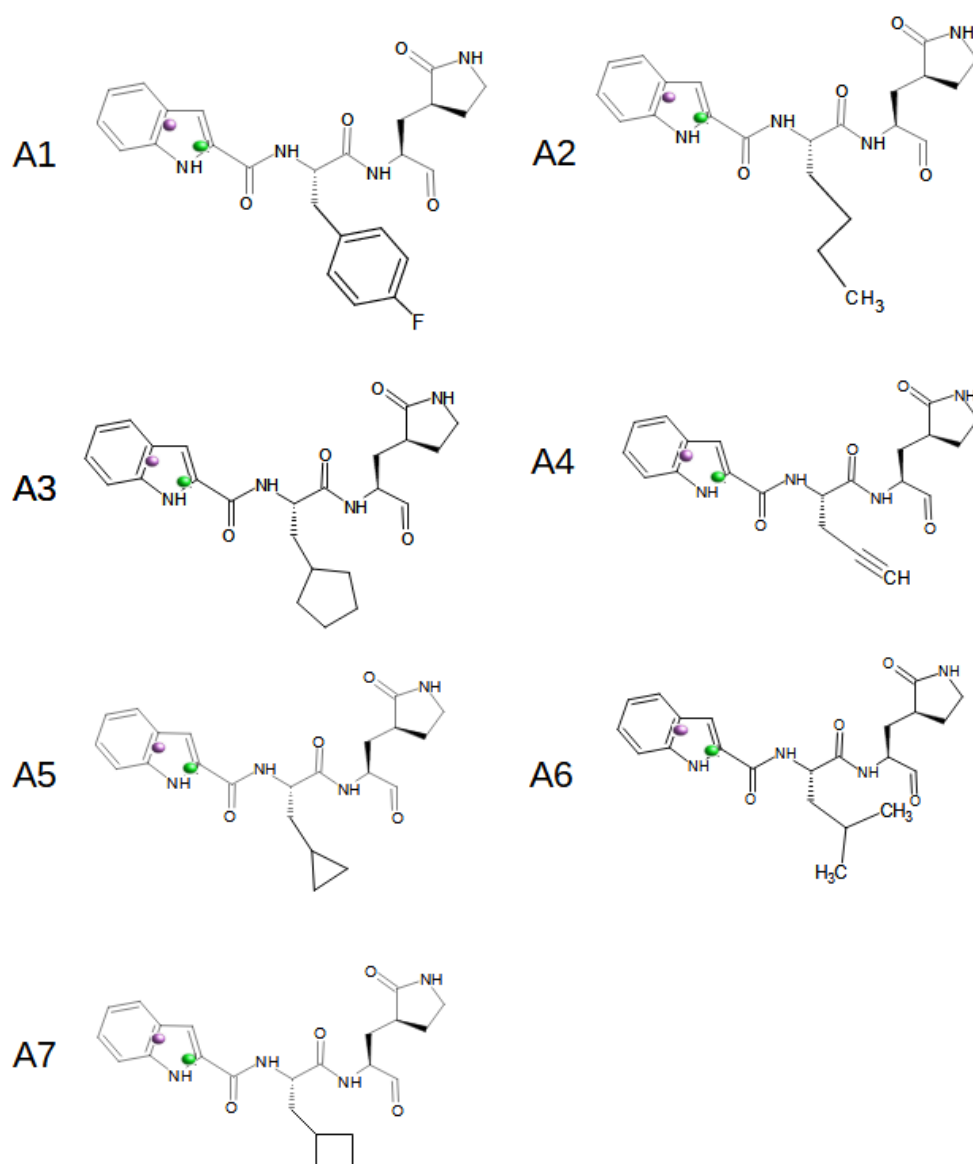
Composto	Estrutura	f^+	f^-	f^0	PEM	Centroids 3D
PAXLOVID						

Fonte: Autoria própria.

APÊNDICE IV

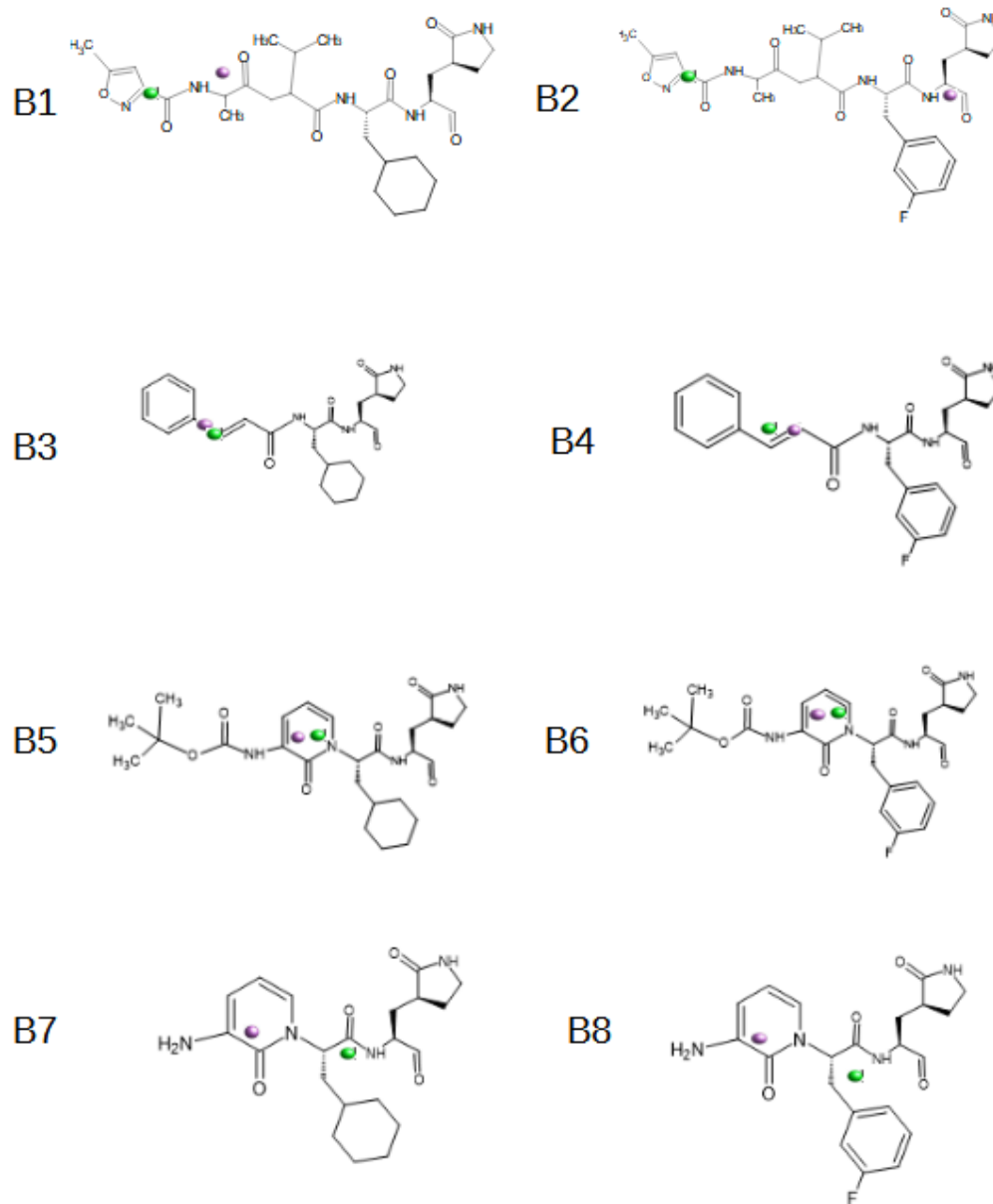
As Figuras A21-A26 apresentam representações em duas dimensões das posições dos centróides associados à f^+ (verde) e f^- (magenta) em relação aos grupos de testagem A, B, C, D e E.

Figura A21. Representação bidimensional da posição dos centróides associados à f^+ (verde) e f^- (magenta) em relação dos compostos do grupo de testagem: A1-A7.



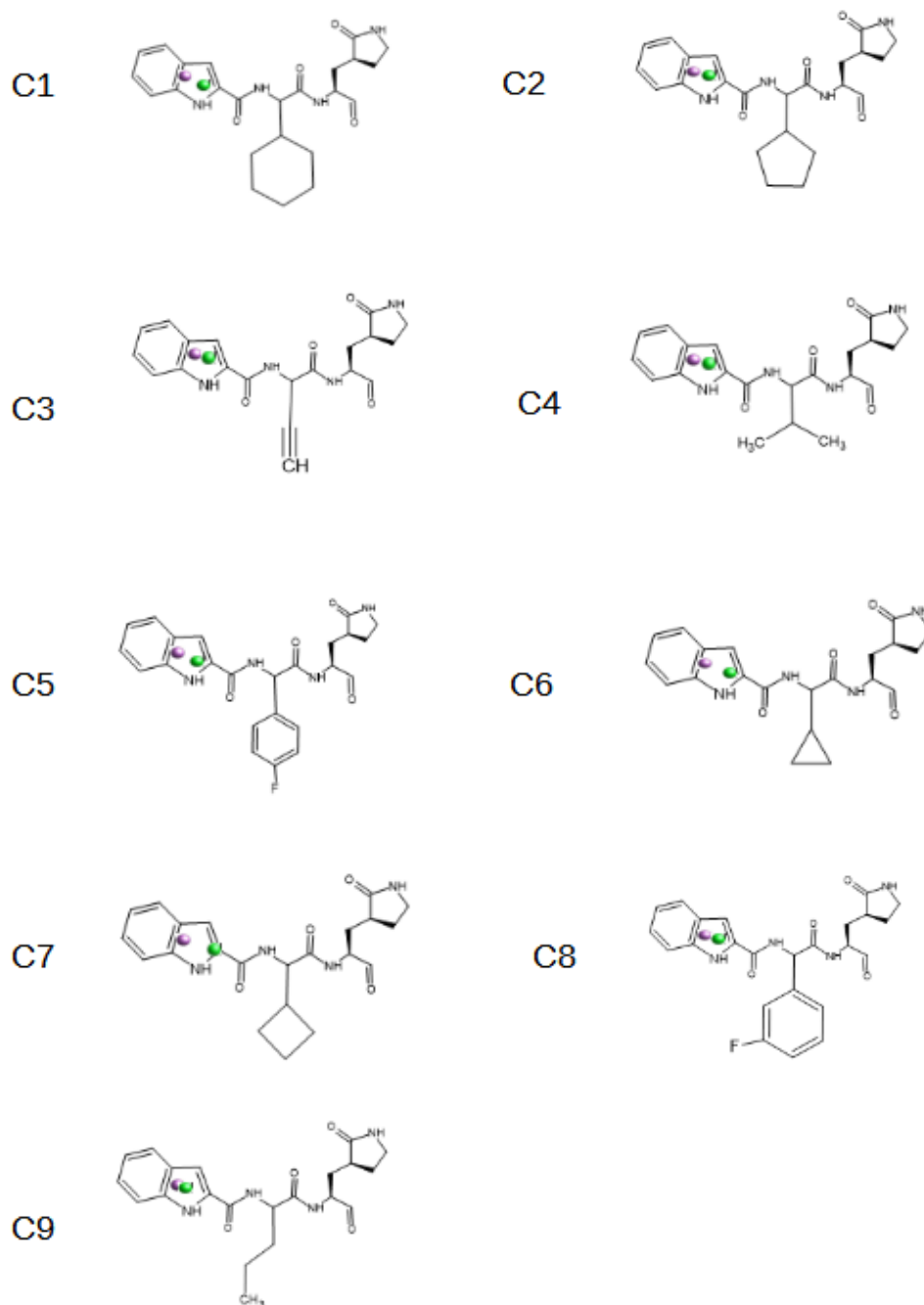
Fonte: Autoria própria.

Figura A22. Representação bidimensional da posição dos centróides associados à f^+ (verde) e f^- (magenta) em relação dos compostos do grupo de testagem: B1-B8.



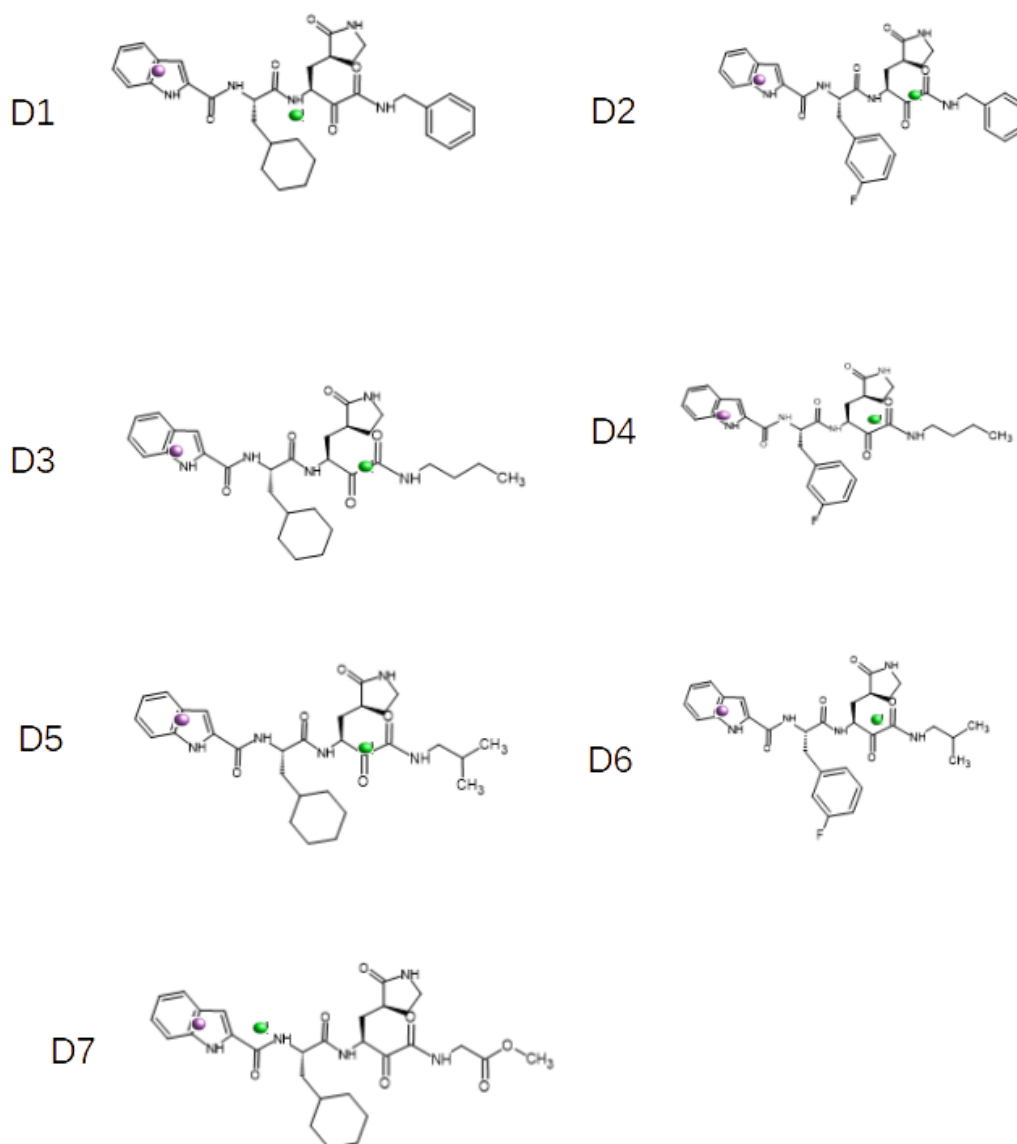
Fonte: Autoria própria.

Figura A23. Representação bidimensional da posição dos centróides associados à f^+ (verde) e f^- (magenta) em relação dos compostos do grupo de testagem: C1-C9.



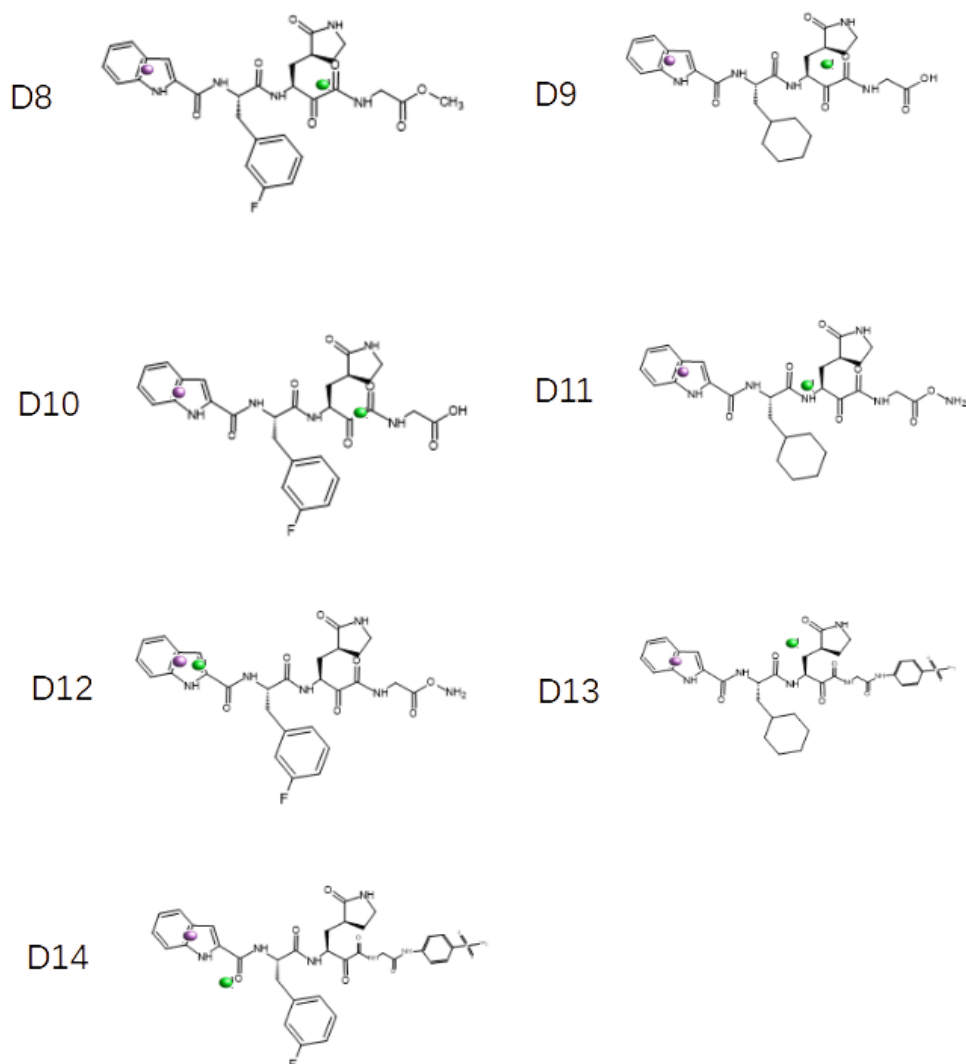
Fonte: Autoria própria.

Figura A24. Representação bidimensional da posição dos centróides associados à f^+ (verde) e f^- (magenta) em relação dos compostos do grupo de testagem: D1-D7.



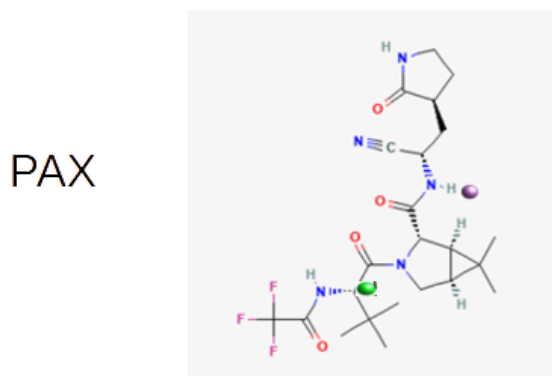
Fonte: Autoria própria.

Figura A25. Representação bidimensional da posição dos centróides associados à f^+ (verde) e f^- (magenta) em relação dos compostos do grupo de testagem: D8-D14.



Fonte: Autoria própria.

Figura A26. Representação bidimensional da posição dos centróides associados à f^+ (verde) e f^- (magenta) em relação ao Paxlovid do grupo E de testagem.



Fonte: Autoria própria.