

Universidade Federal do Rio Grande do Sul  
**Instituto de Química**  
Química Bacharel

Eduardo Ramires Kuhn

Estudo computacional de docagem e dinâmica molecular sobre o mecanismo de  
ação do Molnupiravir sobre o Coronavírus SARS-CoV-2

Porto Alegre  
2023

Eduardo Ramires Kuhn

**Estudo computacional de docagem e dinâmica molecular sobre o mecanismo de ação do Molnupiravir sobre o Coronavírus SARS-CoV-2**

Trabalho de conclusão de curso como requisito parcial à obtenção do título de bacharel em química da Instituto de Química da Universidade Federal do Rio Grande do Sul.

Orientador: Paulo Augusto Netz

Porto Alegre

2023

## CIP - Catalogação na Publicação

Kuhn, Eduardo Ramires

Estudo computacional de docagem e dinâmica molecular sobre o mecanismo de ação do Molnupiravir sobre o Coronavírus SARS-CoV-2 / Eduardo Ramires Kuhn.

-- 2023.

51 f.

Orientador: Paulo Augusto Netz.

Trabalho de conclusão de curso (Graduação) --  
Universidade Federal do Rio Grande do Sul, Instituto de Química, Bacharelado em Química, Porto Alegre, BR-RS, 2023.

1. Molnupiravir. 2. Covid19. 3. RNA polimerase dependente de RNA . 4. Docagem Molecular. 5. Dinâmica Molecular. I. Netz, Paulo Augusto, orient. II. Título.

Eduardo Ramires Kuhn

**Estudo computacional de docagem e dinâmica molecular sobre o mecanismo de ação do Molnupiravir sobre o Coronavírus SARS-CoV-2**

Trabalho de conclusão de curso como requisito parcial à obtenção do título de bacharel em química da Instituto de Química da Universidade Federal do Rio Grande do Sul.

Orientador: Paulo Augusto Netz

**Aprovada em:** Porto Alegre, 9 de abril de 2023.

BANCA EXAMINADORA:

---

Prof. Dr. Paulo Augusto Netz  
Universidade Federal do Rio Grande do Sul

---

Prof. Dr. Jones de Andrade  
Universidade Federal do Rio Grande do Sul

---

Profa. Dra. Silma Alberton Corrêa  
Universidade Federal do Rio Grande do Sul

Dedico este sucinto espaço aos meus mais profundos agradecimentos as pessoas que marcaram minha graduação. Antes de todos, devo mencionar em especial minha mãe, Marli, que me apoiou durante toda minha vida e especialmente durante minha graduação, meu pai Jerri e minha irmã Karine que me auxiliaram sempre que possível. Eterna gratidão aos meus irmãos de vida, Gustavo Toledo e Marcelo Comassetto por sempre me incentivaram a ir em frente. A minha amada, Ana Hofstetter, que conheci durante a graduação e desde então me motiva e incentiva, aguentando todos os momentos mais difíceis do período de graduação comigo.

Gostaria de agradecer em especial ao Igor Silveira que me ajudou muito no início e ao longo da graduação. Também sou muito grato a todos os amigos que fiz na graduação, em especial ao Pedro Acosta, Leonardo Pavan e Julia Lambrecht que contribuíram vivenciando muitos desses momentos comigo.

Ao Paulo Netz, o qual esteve presente em grande parte da minha graduação, sendo meu professor e meu orientador confiando e contribuindo para minha formação hoje. Aos colegas e professores do Grupo de Química Teórica (GQT).

## RESUMO

O novo coronavírus (SARS-CoV-2), sendo o agente que provoca a pandemia CoViD-19, se tornou a maior preocupação da saúde pública. O mecanismo principal de entrada do vírus é através de uma interação inicial da enzima ACE2 (Enzima Conversora da Angiotensina 2) com o SARS-CoV-2, o que possibilita a entrada do vírus e posterior replicação envolvendo duas proteínas não-estruturais RNA polimerase dependente de RNA (RdRp) e Protease Principal ( $M^{PRO}/CL^{PRO}$ ). Há diversas pesquisas em torno de um tratamento eficaz e comprovado, além das vacinas desenvolvidas. Medicamentos como Molnupiravir, uma pró-droga aprovada como medicamento no Reino Unido em 2021, é um derivado nucleosídeo sintético. Em sua forma ativa trifosfato (MTP) atua inibindo a reprodução do vírus por meio de inserção de erros na cópia durante o processo de replicação do RNA. O presente trabalho tem como objetivo aprofundar o estudo em torno do mecanismo de ação do MTP com o seu principal receptor RdRp, expandindo para outro possível receptor que é a Protease Principal ( $M^{PRO}/CL^{PRO}$ ) do SARS-CoV-2. As interações entre as proteínas do SARS-CoV-2 e o MTP foram estudadas mediante cálculos de docagem molecular no AutoDock4, posteriormente analisadas com AutoDockTools (ADT) e PyMOL. Com base nos resultados obtidos, foram conduzidas simulações de dinâmica molecular clássica e obtidos dados de RMSD, RMSF, distância mínima e do centro de massa, ligações de hidrogênio e resíduos na interface de interação. Por fim, foram demonstradas interações favoráveis e permanência explicável do ligante MTP com as proteínas do SARS-CoV-2, indicando que além da atuação com RdRp, é provável atuação do fármaco com a  $M^{PRO}$ .

**Palavras-chave:** Molnupiravir, Covid19, RdRp,  $M^{PRO}$ , Docagem Molecular, Dinâmica Molecular.

## ABSTRACT

The new coronavirus (SARS-CoV-2), being the agent that causes the CoViD-19 pandemic, has become a major public health concern. The main mechanism of entry of the virus is through an initial interaction of the enzyme ACE2 (Angiotensin-Converting Enzyme 2) with SARS-CoV-2, which enables the entry of the virus and subsequent replication involving two non-structural proteins RNA-dependent polymerase RNA (RdRp) and Main Protease ( $M^{PRO}/CL^{PRO}$ ). There are several researches around an effective and proven treatment, in addition to the vaccines developed. Medicines such as Molnupiravir, a pro-drug approved as a medicine in the United Kingdom in 2021, is a synthetic nucleoside derivative. In its active form triphosphate (MTP) acts by inhibiting the reproduction of the virus by inserting errors in the copy during the RNA replication process. The present work aims to improve the understanding of the mechanism of action of MTP with its main receptor RdRp, expanding to another possible receptor that is the Main Protease ( $M^{PRO}/CL^{PRO}$ ) of SARS-CoV-2. Interactions between SARS-CoV-2 proteins and MTP were studied using Molecular Docking calculations in AutoDock4, later analyzed with AutoDockTools (ADT) and PyMOL. Based on the obtained results, classical Molecular Dynamics simulations were performed and RMSD, RMSF, minimum distance and center of mass distance, hydrogen bonds and residues at the interaction interface were obtained. Finally, favorable interactions and explainable permanence of the MTP ligand with the SARS-CoV-2 proteins were demonstrated, indicating that in addition to the action with RdRp, the drug's action with  $M^{PRO}$  is likely.

**Keywords:** Molnupiravir, COVID19, RdRp,  $M^{PRO}$ , Molecular Docking, Molecular Dynamics.

## LISTA DE ILUSTRAÇÕES

Figura 1. Representação das estruturas químicas das formas trifosforilada de Molnupiravir (MTP), Citosina (CTP) e Uracila (UTP).....	13
Figura 2. Representação de interação entre nucleosídeos A e G interagindo com os tautômeros Imino-M e Amino-M, respectivamente. ....	14
Figura 3. Proteína RdRp (NSP12) envolta de outras proteínas não-estruturais (NSP) com o genoma guia positivo em azul e o produto complementar em vermelho em seu sítio de replicação (A), vista lateralmente (B) e a representação de nucleosídeos A, T, G, C, U e o análogo M que gera mutagênese para G ou A.....	14
Figura 4 . Molnupiravir (EIDD-2801) se converte no plasma para EIDD-1931 e depois é convertido para trifosfato (EIDD-1931-5'-Triphosphate), sua forma ativa. ....	20
Figura 5. Esquema da metodologia geral utilizada em ordem de passos junto de breves descrições de pontos presentes. ....	27
Figura 6 . Visão tridimensional do complexo formado pelo ligante MTP e a RdRp através de estudos de Docagem Molecular, sendo (A) a proteína representada em estruturas secundárias, resíduos explícitos no sítio de interação com o MTP representado por superfície roxa. (B) Visão aproximada do sítio ligante com os resíduos de interação explícitos e coloridos de vermelho (contato polar) e azul (não polar) e em destaque com cadeia maior o MTP e resíduos principais.....	31
Figura 7 . Visão tridimensional do complexo formado pelo ligante MTP e a MPRO através de estudos de Docagem Molecular, sendo (A) a proteína representada em estruturas secundárias, resíduos explícitos no sítio de interação com o MTP representado por superfície roxa. (B) Visão aproximada do sítio ligante com os resíduos de interação explícitos e coloridos de vermelho (contato polar) e azul (não polar) e em destaque com cadeia maior o MTP e resíduos principais. ....	32
Figura 8 . Resultados gráficos do cálculo de RMSD para o sistema RdRp e MTP, sendo (A) primeira replicata e (B) segunda replicata. ....	33
Figura 9. Resultados gráficos do cálculo de distância mínima e entre os centros de massa para o sistema RdRp e MTP, sendo (A) primeira replicata e (B) segunda replicata.....	34
Figura 10. Resultados gráficos das ligações de hidrogênio para o sistema RdRp e MTP, sendo (A) primeira replicata e (B) segunda replicata. ....	35



Figura 11. Resultados gráficos do cálculo de RMSF para o sistema RdRp e MTP, sendo (A) primeira replicata e (B) segunda replicata. ....	36
Figura 12. Visão tridimensional do complexo formado pelo ligante MTP e a RdRp através de estudos de DM no tempo de 100ns, sendo (A) primeira replicata e a (C) segunda replicata com a proteína representada em estruturas secundárias, resíduos no sítio de interação com o MTP representado por superfície roxa. (B) primeira replicata e a (D) segunda replicata com visão aproximada do sítio ligante com os resíduos de interação explícitos e coloridos de vermelho (contato polar) e azul (não polar) e em destaque com cadeia maior o MTP e resíduos principais. ....	37
Figura 13. Visão tridimensional do complexo formado pelo ligante MTP e a RdRp através de estudos de DM no tempo de 200ns, sendo (A) primeira replicata e a (C) segunda replicata com a proteína representada em estruturas secundárias, resíduos no sítio de interação com o MTP representado por superfície roxa. (B) primeira replicata e a (D) segunda replicata com visão aproximada do sítio ligante com os resíduos de interação explícitos e coloridos de vermelho (contato polar) e azul (não polar) e em destaque com cadeia maior o MTP e resíduos principais. ....	38
Figura 14. Resultados gráficos do cálculo de RMSD para o sistema MPRO e MTP, sendo (A) primeira replicata e (B) segunda replicata. ....	39
Figura 15. Resultados gráficos do cálculo de distância mínima e entre os centros de massa para o sistema MPRO e MTP, sendo (A) primeira replicata e (B) segunda replicata. ....	40
Figura 16. Resultados gráficos das ligações de hidrogênio para o sistema MPRO e MTP, sendo (A) primeira replicata e (B) segunda replicata. ....	40
Figura 17. Resultados gráficos do cálculo de RMSF para o sistema MPRO e MTP, sendo (A) primeira replicata e (B) segunda replicata. ....	41
Figura 18. Visão tridimensional do complexo formado pelo ligante MTP e a MPRO através de estudos de DM no tempo de 100ns, sendo (A) primeira replicata e a (C) segunda replicata com a proteína representada em estruturas secundárias, resíduos no sítio de interação com o MTP representado por superfície roxa. (B) primeira replicata e a (D) segunda replicata com visão aproximada do sítio ligante com os resíduos de interação explícitos e coloridos de vermelho (contato polar) e azul (não polar) e em destaque com cadeia maior o MTP e resíduos principais. ....	42
Figura 19. Visão tridimensional do complexo formado pelo ligante MTP e a MPRO através de estudos de DM no tempo de 200ns para a primeira replicata e 175ns na	

segunda, sendo (A) primeira replicata e a (C) segunda replicata com a proteína representada em estruturas secundárias, resíduos no sítio de interação com o MTP representado por superfície roxa. (B) primeira replicata e a (D) segunda replicata com visão aproximada do sítio ligante com os resíduos de interação explícitos e coloridos de vermelho (contato polar) e azul (não polar) e em destaque com cadeia maior o MTP e resíduos principais. ....43

## LISTA DE ABREVIATURAS E SIGLAS

A	Adenosina
ALA	Alanina
LGA	Algoritmo Genético Lamarckiano
ARG	Arginina
ASN	Asparagina
ASX	Asparagina ou Aspartato
ASP	Aspartato (Ácido Aspártico)
ADT	AutoDockTools
CESUP	Centro Nacional De Super Computação
CYS	Cisteína
CTP	Cistidina Trifosforilada
CoViD-19	Coronavírus
RMSD	Desvio Quadrático Médio
DM	Dinâmica Molecular
ACE2	Enzima Conversora da Angiotensina 2
PHE	Fenilalanina
-gRNA	Genoma Guia De RNA Negativo
+gRNA	Genoma Guia De RNA Positivo
GLY	Glicina
GLU	Glutamato (Ácido Glutâmico)
GLN	Glutamina (Glutamida)
GLX	Glutamina ou Glutamato
G	Guanina
HIS	Histidina
ILE	Isoleucina
GQT	Laboratório do Grupo Química Teórica da UFRGS
LEU	Leucina
LYS	Lisina
MET	Metionina
MMP	Molnupiravir Monofosfatado
MTP	Molnupiravir Trifosforilado

NTP	Nucleosídeos Trifosforilados
PRO	PROLINA
M <sup>PRO</sup>	Protease Principal ou Main Protease
CL <sup>PRO</sup>	Protease Principal ou Main Protease
RMSF	Raiz da Flutuação Quadrada Média
RdRp	RNA polimerase dependente de RNA
SER	Serina
SARS-CoV-2	Síndrome Respiratória Aguda Grave Coronavírus 2
SPME	Smooth Particle Mesh Ewald
DFT	Teoria Funcional Da Densidade
TYR	Tirosina
THR	Treonina
TRP	Triptofano (Triptofana)
UTP	Uracila Trifosforilada
VAL	Valina

## SUMÁRIO

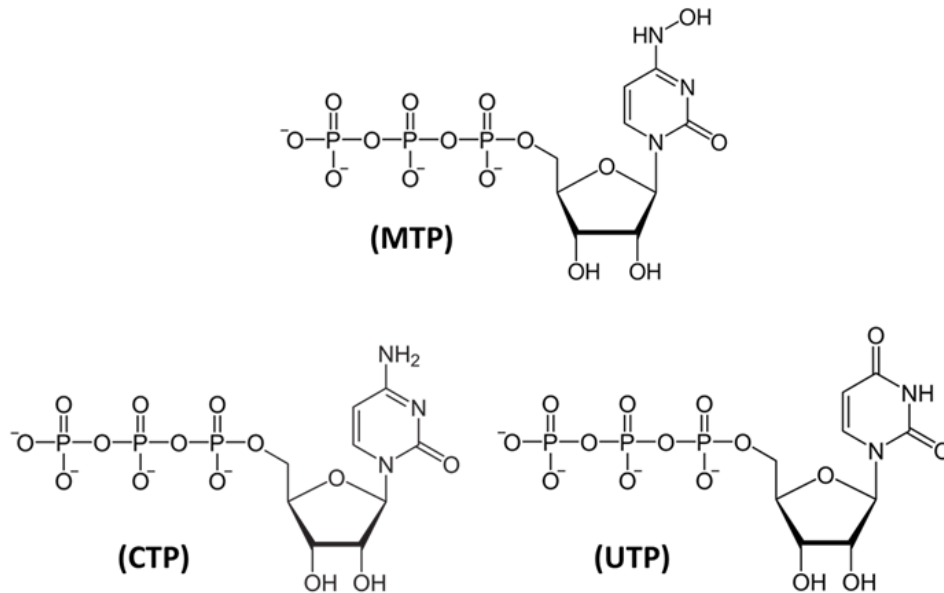
<b>1 INTRODUÇÃO</b> .....	<b>12</b>
<b>2 OBJETIVOS</b> .....	<b>16</b>
2.1 OBJETIVO GERAL .....	16
2.2 OBJETIVOS ESPECÍFICOS .....	16
<b>3 FUNDAMENTAÇÃO TEÓRICA</b> .....	<b>17</b>
3.1 CORONAVÍRUS SARS-COV-2 E SEUS POTENCIAIS ALVOS .....	17
3.2 PLANEJAMENTO RACIONAL E REPOSICIONAMENTO DE FÁRMACOS .....	18
<b>3.2.1 Molnupiravir</b> .....	<b>19</b>
3.3 QUÍMICA TEÓRICA E COMPUTACIONAL.....	21
<b>3.3.1 Docagem Molecular</b> .....	<b>21</b>
<b>3.3.2 Dinâmica Molecular</b> .....	<b>23</b>
<b>3.3.2.1 Campos de Força</b> .....	<b>24</b>
<b>3.3.2.2 Condições Periódicas de Contorno e Ensemble</b> .....	<b>25</b>
<b>4 METODOLOGIA</b> .....	<b>26</b>
4.1 BANCO DE DADOS .....	26
4.2 DOCAGEM MOLECULAR.....	27
4.3 DINÂMICA MOLECULAR.....	28
4.4 ANÁLISE DOS RESULTADOS .....	30
<b>5 RESULTADOS E DISCUSSÕES</b> .....	<b>30</b>
5.1 DOCAGEM MOLECULAR.....	30
5.2 DINÂMICA MOLECULAR.....	32
<b>5.2.1 RNA polimerase dependente de RNA (RdRp) e MTP</b> .....	<b>33</b>
<b>5.2.2 Protease Principal (M<sup>PRO</sup>) e MTP</b> .....	<b>39</b>
<b>6 CONSIDERAÇÕES FINAIS</b> .....	<b>44</b>
<b>REFERÊNCIAS</b> .....	<b>45</b>

## 1 INTRODUÇÃO

O novo coronavírus, sendo o agente que provoca a pandemia de CoViD-19 ou da síndrome respiratória aguda grave coronavírus 2 (SARS-CoV-2), se tornou a maior preocupação da saúde pública. O vírus provavelmente originário de morcegos foi transmitido aos humanos em Wuhan, província de Hubei, China, em dezembro de 2019, causando sintomas como: febre, tosse, dor de garganta, falta de ar, cansaço e mal-estar. Em casos mais graves, pode evoluir para pneumonia e disfunção de múltiplos órgãos. Devido à alta taxa de transmissão do vírus, logo se tornou uma preocupação mundial obrigando muitos países a determinar estado de quarentena e paralisar boa parte do cotidiano não-essencial (ANDERSEN et al., 2020; GUO et al., 2020; WANG et al., 2020). De acordo com os relatórios estatísticos mais recentes do WorldMeters, o coronavírus já afetou mais de 680 milhões de pessoas no mundo inteiro, causando mais de 6 milhões de mortes (“COVID - Coronavirus Statistics - Worldometer”).

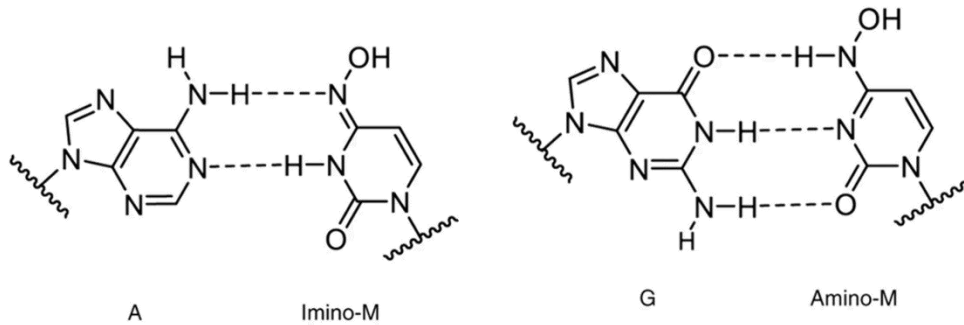
O mecanismo principal de entrada do vírus é através de uma interação inicial da enzima ACE2 (Enzima Conversora da Angiotensina 2) com o SARS-CoV-2, o que possibilita a entrada do vírus, liberando 4 proteínas estruturais (Spike (S), envelope (E), membrana (M) e nucleocapsídeo (N)), 16 proteínas não-estruturais (NSP1-NSP16) e diversas proteínas acessórias. Uma vez liberado o material genético do vírus na célula, inicia-se a sua replicação envolvendo as proteínas não-estruturais RNA polimerase dependente de RNA (RdRp) e Main Protease ( $M^{PRO}/CL^{PRO}$ ). Há diversas pesquisas em torno de um tratamento eficaz e comprovado, além das vacinas desenvolvidas, medicamentos como Remdesivir, Sotrovimabe, Baricitinibe, Evusheld, Paxlovid e Molnupiravir foram aprovados pela Anvisa uso emergencial (“Anvisa aprova venda do medicamento Lagevrio (molnupiravir) em farmácias — Agência Nacional de Vigilância Sanitária - Anvisa”; SHANG et al., 2020; WU et al., 2020).

Figura 1. Representação das estruturas químicas das formas trifosforilada de Molnupiravir (MTP), Citosina (CTP) e Uracila (UTP).



Dentre elas, o Molnupiravir é uma pró-droga aprovada como medicamento no Reino Unido em 2021 (SYED, 2022). É um análogo de nucleosídeo sintético que em sua forma trifosforilada (MTP) busca atuar na proteína RdRp, seu principal alvo, por meio de inserção na replicação do RNA viral (SINGH et al., 2021). Devido a sua semelhança com outros nucleosídeos trifosforilados (NTP) (Figura 1), pode mimetizar tanto a citosina trifosforilada (CTP) em sua forma amino, quanto uracila trifosforilada (UTP) graças ao seu tautomerismo (Figura 2). Assim, quando inserido no lugar de uma dessas bases, pode levar ao erro na replicação da fita invertendo qual nucleosídeo é inserido na fita de RNA viral, como mostrado na Figura 3 (C). O mecanismo de revisão dos nucleosídeos incorporados no RNA replicado impede que haja erros, porém o Molnupiravir após a inserção no RNA parece não ser afetado, permitindo que o mecanismo de replicação da fita de RNA viral continue, uma vez que muitas drogas são descartadas após a inserção na cadeia por mecanismo de revisão do próprio vírus (SINGH et al., 2021).

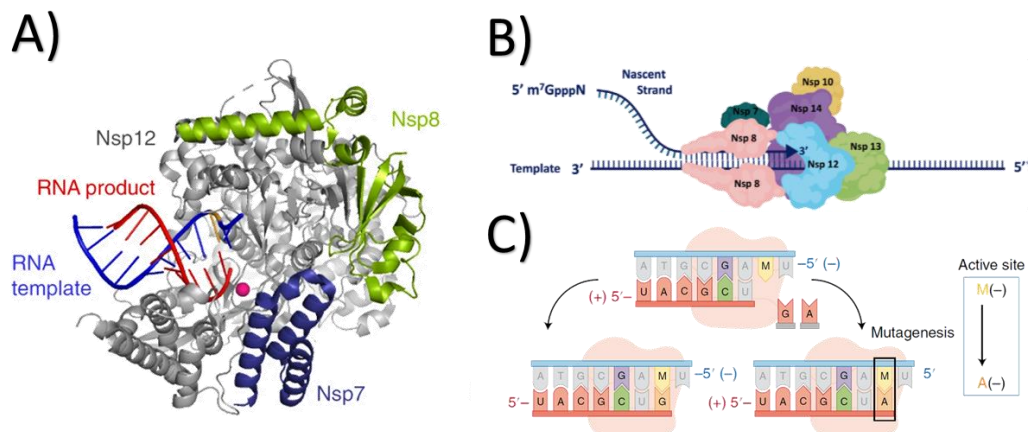
Figura 2. Representação de interação entre nucleosídeos A e G interagindo com os tautômeros Imino-M e Amino-M, respectivamente.



Fonte: (KABINGER et al., 2021)

Além do seu principal alvo RdRp, sabe-se que é provável a interação com outros alvos no interior da célula, podendo causar interações medicamentosas (por exemplo a ação no mesmo alvo de inibição). Dessa forma, o estudo do MTP com outras proteínas envolvidas no processo de replicação do vírus após a entrada na célula é válido. A interação com a Protease Principal ( $M^{PRO}$ ) do SARS-CoV-2 é um exemplo muito explorado na literatura com a forma monofosfatada (MMP) (CALLIGARI et al., 2020).

Figura 3. Proteína RdRp (NSP12) envolta de outras proteínas não-estruturais (NSP) com o genoma guia positivo em azul e o produto complementar em vermelho em seu sítio de replicação (A), vista lateralmente (B) e a representação de nucleosídeos A, T, G, C, U e o análogo M que gera mutagênese para G ou A.



Fonte: (KABINGER et al., 2021; MALONE; CAMPBELL, 2021; ROBSON et al., 2020)



O presente trabalho tem como objetivo aprofundar o estudo em torno do mecanismo de ação do Molnupiravir em sua forma ativa trifosforilada (MTP) com seu principal receptor RNA polimerase dependente de RNA (RdRp) e expandindo para outro possível receptor, a Protease Principal ( $M^{PRO}/CL^{PRO}$ ), utilizando de métodos como a Docagem Molecular e a Dinâmica Molecular (KUSHWAHA et al., 2021).

As proteínas RdRp e  $M^{PRO}$  do SARS-CoV-2 foram obtidas pelo banco de dados *Protein Data Bank* (PDB) gerenciada pela *Research Collaboratory for Structural Bioinformatics* (RCSB) com os identificadores 6m71 e 6lu7, respectivamente. Já a estrutura do molnupiravir trifosforilado (MTP) foi construída e posteriormente otimizada com o servidor online *Automatic Topology Builder* (ATB) (MALDE et al., 2011). As interações entre as proteínas do SARS-CoV-2 e o MTP foram estudadas mediante cálculos de Docagem Molecular no AutoDock4 (ALLOUCHE, 2012; “AutoDock”) e analisadas com AutoDockTools (ADT) (MORRIS et al., 2009). Com base nos resultados obtidos, foram conduzidas simulações de Dinâmica Molecular (DM) clássica no pacote GROMACS (ABRAHAM et al., 2015) e obtidos dados de Desvio Quadrático Médio (*Root Mean Square Deviation*, RMSD), Raiz da Flutuação Quadrada Média (*Root Mean Square Fluctuation*, RMSF), distâncias entre os centros de massa do ligante e da proteína junto com a distância mínima entre eles, ligações de hidrogênio e resíduos na face de interação que foram utilizados na demonstração de interações e na investigação da permanência do ligante no sítio ativo das enzimas do SARS-CoV-2 estudadas.

## 2 OBJETIVOS

### 2.1 OBJETIVO GERAL

Aprofundar o estudo em torno do mecanismo de ação do Molnupiravir em sua forma trifosforilada (MTP) com seu principal receptor RNA polimerase dependente de RNA (RdRp) e para outro possível receptor Protease Principal ( $M^{PRO}$ ) do SARS-CoV-2, utilizando os métodos computacionais: Docagem Molecular e Dinâmica Molecular (DM).

### 2.2 OBJETIVOS ESPECÍFICOS

Investigar por Docagem Molecular os sistemas RdRp+MTP e  $M^{PRO}$ +MTP.

A partir do resultado de Docagem, simular DM dos sistemas investigando a permanência e as interações no sítio de ligação.

### 3 FUNDAMENTAÇÃO TEÓRICA

#### 3.1 CORONAVÍRUS SARS-COV-2 E SEUS POTENCIAIS ALVOS

O novo coronavírus SARS-CoV-2 (*Severe acute respiratory syndrome coronavirus-2* ou síndrome respiratória aguda grave coronavírus-2) foi detectado em 2019 tendo um impacto maior em relação aos outros coronavírus antes conhecidos (SARS-CoV-1, MERS CoV) (SINGHAL, 2020). Cerca de 77% das sequências de proteínas do SARS-CoV-2 também estão presentes no SARS-CoV-1, assim grande parte do acervo de pesquisas relacionadas ao SARS-CoV-1 são relevantes no estudo e desenvolvimento de fármacos para a Covid19 (CERAOLO; GIORGI, 2020). Dessa forma, a compreensão das proteínas presentes nesses vírus permite uma abordagem racional no desenvolvimento de fármacos mais eficazes a partir do conhecimento estrutural dos envolvidos. (CALLIGARI et al., 2020; HILGENFELD, 2014).

As proteínas do SARS-CoV-1 consistem em duas grandes poliproteínas (ORF1a e ORF1ab) que se clivam proteoliticamente para formar 16 proteínas não-estruturais (NSP1-NSP16), quatro proteínas estruturais (spike (S), envelope (E), membrana (M) e nucleocapsídeo (N)) e proteínas acessórias (algumas não presentes no SARS-CoV-2) (SHANG et al., 2020). Dentre as proteínas estruturais, a spike é tida como principal alvo para o desenvolvimento de um potencial droga, enquanto que para as não-estruturais: a protease do tipo 3C ou Protease Principal ( $M^{PRO}$ , NSP5), a protease semelhante à papaína ( $PL^{PRO}$ , NSP3) e a RNA polimerase dependente de RNA viral (RdRp, NSP12) que são as mais investigadas como potenciais alvos (RAJ, 2021; SU et al., 2021).

A entrada na célula hospedeira é mediada pela proteína S do vírus, que é responsável pelo reconhecimento do receptor e no processo de fusão da membrana via interação com a enzima conversora de angiotensina 2 (ACE2) (SHANG et al., 2020). Após a entrada na célula hospedeira, o material genético será liberado e as poliproteínas serão processadas. A  $M^{PRO}$ , que possui uma díade catalítica Cys145-His41 no sítio ativo, exercendo papel fundamental na replicação e maturação do vírus (HU et al., 2022; KUSHWAHA et al., 2021; MURAMATSU et al., 2016). No decorrer da replicação viral, a RdRp é utilizada na replicação e transcrição do genoma guia de RNA positivo (+gRNA) do vírus para seu complementar negativo (-gRNA), o qual passa então a gerar transcrições adicionais (Figura 3). Esse processo é alvo para

antivirais análogos de nucleosídeos que interferem na replicação substituindo os trifosfatos de nucleosídeos (NTP) apropriados para a formação da ligação fosfodiéster, estendendo a extremidade 3' (SOLA et al., 2015; XU et al., 2003) e causando algum efeito diferente da codificação original, podendo esse ser a mutagêneses que podem levar à chamada “catástrofe do erro” (MOLNUPIRAVIR) ou a própria terminação de cadeia também (REMDESIVIR e FAVIPIRAVIR) (SINGH et al., 2021).

O sítio ativo da RdRp é constituído de resíduos chaves(WU et al., 2020) tais como:

ASP-623: reconhecimento do NTP;

SER-682 e THR-687: reconhecimento do iniciador;

SER-682, THR-687 e ASN-691: ajuda na seleção do NTP;

SER-759: ligação do nucleotídeo iniciador;

CYS-813 e SER-814: posicionamento do nucleotídeo;

LYS-545, LYS-551 e ARG-553: ligação e posicionamento do NTP.

Já o sítio ativo da M<sup>PRO</sup> é definido por vezes em 5 sub sítios de interação S1-5, sendo o S1 definido por PHE-140, GLY-143, SER-144, CYS-145, HIS-163, GLU-166 e HIS172 e o S2 por THR-25, HIS-41 e CYS-145(HU et al., 2022). Apenas S1 e S2 são destacados por conterem HIS41 e CYS-145, resíduos principais do sítio catalítico.

### 3.2 PLANEJAMENTO RACIONAL E REPOSICIONAMENTO DE FÁRMACOS

A pesquisa e desenvolvimento de fármacos é uma necessidade antiga para a sociedade, porém esse processo é altamente custoso e também demorado devido as etapas necessárias para comprovação de sua eficácia, as quais podem levar até 15 anos (LOMBARDINO; LOWE, 2004; OU-YANG et al., 2012; PILLAIYAR et al., 2020; SILVERMAN; HOLLADAY, 2004). Porém, o avanço de técnicas de resolução das estruturas químicas, tais como Difração de Raios-X e Ressonância Magnética Nuclear, permitiu um maior conhecimento estrutural e conformacional de macromoléculas biológicas, contribuindo no surgimento do planejamento racional de fármacos (PELLEGRINI; STÖHR, 2003; SCHLICHTING; MIAO, 2012).

O processo de desenho racional começa pelo estudo do alvo, entendendo a relação de estrutura e atividade com o possível fármaco (composto natural ou

sintético) através da verificação de parâmetros como lipofilicidade, possibilidade de ligações de hidrogênio (quantos doadores ou aceptores possui) e o peso molecular. Os métodos computacionais facilitam muito a pesquisa desses parâmetros, fornecendo ferramentas para o processamento de estruturas cristalográficas (ou modeladas parcialmente também) de determinado receptor, induzindo o encaixe de ligantes no sítio ativo de interesse e obtendo informações mais precisas sobre o sistema. Porém, nem sempre é possível isolar a macromolécula de interesse com os ligantes de forma fácil ou completa (SU et al., 2021).

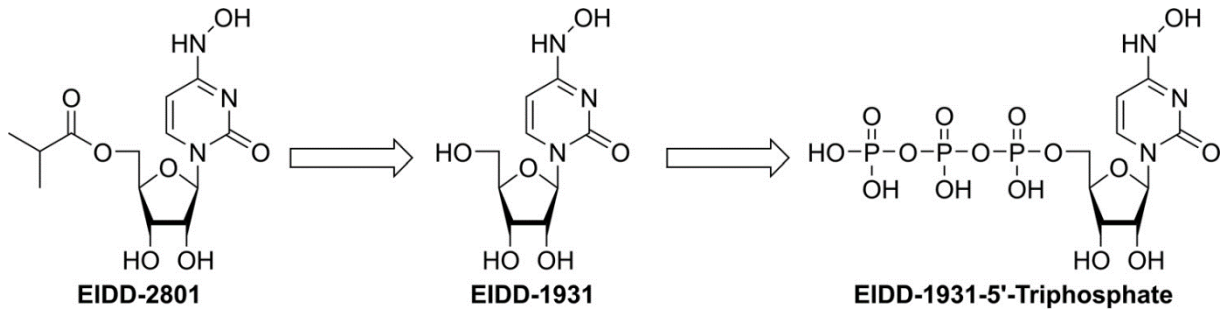
A busca de modelos com potencial interação biológica se dá muitas vezes sobre a pesquisa da relação estrutura e atividade, definindo muitas vezes uma estrutura base, chamada de *scaffold*. Essa linha de raciocínio gerou também os estudos de reposicionamento de fármacos, investigando alvos que tenham maior afinidade ou atividade em comparação com sua finalidade originária. Os fármacos visados para reposicionamento, em geral, já são bem conhecidos em questão de toxicidade, avançado em testes pré-clínicos e também aprovados, facilitando e reduzindo muito o tempo de desenvolvimento caso bem sucedido (PILLAIYAR et al., 2020).

A utilização de modelos tridimensionais no desenho de fármacos baseado na estrutura de alvos biológicos (*structure-based drug design*, SBDD), tem ganhado popularidade, mas ainda existe um custo computacional para gerar dados de simulação com diferentes métodos dependente das dimensões e complexidade do sistema e do programa. Nessa abordagem, já existem bancos de dados bem difundidos com diversas estruturas de potenciais ligantes e macromoléculas, sendo necessário apenas o tratamento de dados de acordo com a pesquisa.

### 3.2.1 Molnupiravir

O molnupiravir é um pró-fármaco de éster isopropílico que é clivado no plasma para um análogo de ribonucleosídeo de pirimidina ativo  $\beta$ -D-N4-hidroxicitidina (NHC) ou EIDD-1931, esse por sua vez é distribuído pelos tecidos e incorporado nas células sendo convertido a sua forma NHC trifosfato (MTP) (Figura 4), tendo como alvo a RdRp para ser codificada competitivamente com os nucleotídeos trifosfato de citidina (CTP) e uridina (UTP).

Figura 4 . Molnupiravir (EIDD-2801) se converte no plasma para EIDD-1931 e depois é convertido para trifosfato (EIDD-1931-5'-Triphosphate), sua forma ativa.



Fonte: (PAINTER et al., 2021)

A presença do trifosfato de molnupiravir junto de outros nucleosídeos é incorporado na síntese do RNA genômico de fita negativa (-gRNA) pela transcrição do RNA genômico de fita positiva (+ gRNA), que em uma segunda etapa irá transcrever um novo +gRNA com mutações na transcrição do RNA levando à “catástrofe do erro”. A escolha de qual base o molnupiravir compete é devido ao seu tautomerismo (LEŚ; ADAMOWICZ; RODE, 1993) na sua forma amino, MTP mimetiza Citosina (C) e pareia com Guanina (G), enquanto que na forma imino, mimetiza Uracila (U) e pareia com Adenosina (A) (Figura 2).

Esse fármaco inicialmente foi desenvolvido pela universidade de Emory (EUA) com foco no tratamento de infecções por vírus, como influenza, mas com o avanço da epidemia o foco do grupo se voltou ao reposicionamento para o alvo RdRp da SARS-CoV-2, o que mostrou muitos resultados positivos no tratamento da Covid19. Após a incorporação do molnupiravir como monofosfato (M) no RNA nascente, não há impedimento da continuidade de extensão do RNA, sendo conduzido ao erro na transcrição.

O fato de que esse fármaco age mimetizando pares de bases presentes no DNA também levantou questões sobre a segurança das nossas células não infectadas (STICHER et al., 2020; ZHOU et al., 2021), mas a administração do medicamento deve ser de forma oral e por curto período de tempo (5 dias) apresentou resultados efetivos sem danos consideráveis impeditivos nas fases pré clínicas 1, 2 e 3(SINGH et al., 2021).

### 3.3 QUÍMICA TEÓRICA E COMPUTACIONAL

O desenvolvimento de ferramentas computacionais está presente em todas as áreas de conhecimento, inclusive as mais tradicionalmente experimentais, como a Química. Dentro dela, a base de conhecimento e desenvolvimento dessas ferramentas que utilizam metodologias matemáticas e modelos físicos capazes de descrever propriedades químicas é chamada de Química Teórica, enquanto que a área que emprega os métodos computacionais disponíveis nos estudos dos mais variados sistemas químicos é chamada de Química Computacional.

Diversas metodologias foram e seguem sendo desenvolvidas a fim de entender e elucidar os sistemas químicos. Dessa forma, informações de propriedades farmacodinâmicas e farmacocinéticas podem ser extraídas e avaliadas através dessas ferramentas. No presente trabalho serão abordadas duas metodologias principais, a Docagem Molecular (ou Docking Molecular) e a Dinâmica Molecular (DM), muito aplicadas nos estudos de sistemas bioquímicos, tais como interação receptor-substrato, avaliação conformacional de macromoléculas, avaliação energética e na avaliação de biomoléculas e substratos específicos. A DM é complementar ao estudo estático de ancoragem dos ligantes por incluir os graus de liberdade do sistema em função do tempo, gerando diversas informações valiosas para definir mecanismos de interação, bem como sua permanência no sistema.

#### 3.3.1 Docagem Molecular

Docagem Molecular é um método computacional baseado em cálculo de ancoragem entre moléculas: geralmente entre uma molécula pequena chamada de ligante, e um alvo macromolecular chamado de receptor. Essa metodologia é amplamente utilizada no planejamento racional de fármacos por ter a capacidade de prever a conformação e a orientação de pequenas moléculas no sítio de ligação em alvos apropriados, formando um complexo receptor-ligante.

A simulação de ancoragem é composta por basicamente dois métodos principais: primeiramente um método de busca conformacional através do espaço disponível demonstrando vários modos de ligações potenciais, seguido por um método de avaliação energética de cada uma das conformações através de uma determinada função de escore ou pontuação. Os algoritmos de busca conformacional

exploram a flexibilidade do ligante usando métodos de busca sistemática e/ou estocástica. Os métodos de busca sistemática procuram uma mudança gradativa de conformações partindo de pequenas variações nos parâmetros estruturais, investigando após vários ciclos de avaliação a convergência para os mínimos de energia obtidos, os quais correspondem ao modo de ligação mais provável. Já nos métodos estocásticos, a busca é realizada com modificações aleatórias dos parâmetros estruturais do ligante promovendo uma ampla cobertura ao gerar conjuntos de conformações e aumentando a chance de aproximar do mínimo global.

Os algoritmos genéticos, um exemplo de algoritmos estocásticos, são baseados na teoria de evolução das espécies. Neles é gerada uma primeira população aleatória de conformações, avaliadas energeticamente, e mantidas as conformações com melhores interações como progenitores de uma nova população aleatória baseada em mutação e/ou crossing-over de suas características. Esse processo é realizado sucessivamente até um limite definido previamente nos parâmetros da ferramenta. O Algoritmo Genético Lamarckiano (LGA) busca simular um comportamento análogo à evolução biológica, iniciando com uma população de conformações criadas e, após gerações sucessivas de mutações: trocas de parâmetros conformacionais, competindo e selecionando os de menor energia livre de ligação (MORRIS et al., 1998).

As funções de pontuação nos programas de Docagem Molecular servem como parâmetro de ordenamento das conformações baseadas em sua energia de ligação e modo de ligação do complexo proteína-ligante. A função também permite a formação de grupamentos, chamados clusters, de conformações dispostas da mesma região e pequenas alterações energéticas. Dessa forma, a investigação de elementos de ligação favoráveis ou não podem ser identificados e explorados no desenvolvimento do planejamento de um novo fármaco com potencial atividade biológica.

Na definição de parâmetros para a simulação de Docagem do sistema, em geral, é mantida a macromolécula rígida (ou parcialmente rígida) e o ligante torcionalmente flexível, sendo então limitada pela conformação inicial da macromolécula. Assim, muitos estudos combinam as técnicas de Docagem com as de DM, fornecendo maiores informações da flexibilidade da macromolécula e permanência do ligante no sítio estudado do sistema em avaliação.



### 3.3.2 Dinâmica Molecular

A simulação do comportamento de um dado sistema pode ser compreendido pelo comportamento das componentes desse sistema, sendo a Dinâmica Molecular (DM) uma ferramenta computacional que nos permite este estudo por meio de simulação. Essa ferramenta aplica as equações de movimento de Newton da mecânica clássica, fornecendo informações sobre a trajetória e evolução temporal ao nível atômico do sistema químico preparado.

Nessa técnica, parte-se de uma configuração inicial pré-determinada do sistema. A posição dos átomos fornecem informação suficiente para o cálculo das forças que atuam sobre cada partícula, sendo essas intra e intermoleculares. Os programas que realizam a simulação por DM computam para um conjunto de coordenadas as acelerações resultantes das forças sobre cada átomo, gerando novas velocidades e consequentemente novas coordenadas após um passo de tempo pré-determinado. Realizando sucessivos cálculos com tempos regulares é obtida uma trajetória do sistema representando as coordenadas do sistema em função do tempo. Muitas propriedades macroscópicas podem ser extraídas a partir das médias das trajetórias através da Mecânica Estatística, sendo para isso necessárias trajetórias com intervalos de tempo suficientemente longos na escala atômica, cerca de  $10^{-9}$  a  $10^{-6}$  segundos de tempo real suficientemente longo para os sistemas de proteína-ligante.

As energias envolvidas nos processos de simulações de DM ocorrem em uma faixa de ordem de  $k_B T$  ( $k_B$ = constante de Boltzmann e  $T$ =temperatura em Kelvin, faixa de 100 a 800 K), nesse intervalo são bem avaliadas as interações entre átomos, moléculas ou íons e eventuais mudanças conformacionais. Sendo assim, processos como reações químicas, uma vez que há ruptura de ligações e menção explícita à estrutura eletrônica não podem ser descritos pela DM convencional. No entanto hoje é possível com métodos híbridos envolvendo cálculos clássicos e quânticos descrever sistemas mais complexos e reacionais. Apesar disso, no presente trabalho será abordada a DM essencialmente a clássica, sendo desconsiderados fenômenos quânticos, como quebra e formação de ligações químicas e deslocalização ou interação de orbitais eletrônicos.

A DM possui diversidade em aplicações, desde os mais simples como propriedades de gases ou líquidos puros, passando por moléculas e macromoléculas

em solventes, até sistemas complexos como presença de cofatores, biomembranas e proteínas.

A inicialização de um sistema para o estudo de DM necessita da configuração inicial do sistema com a geometria da “caixa” definida. Nela são aplicadas condições periódicas de contorno e a convenção de imagem mínima ou raio de corte esférico. Geralmente em sistemas bioquímicos (em função da conformação experimental apenas da proteína geralmente disponível) existem etapas adicionais prévias de solvatação, inserção de íons e uma minimização de energia com restrição de movimento para relaxação das moléculas de solvente inseridas. O cálculo das forças da DM leva em consideração a distância de cada átomo  $i$  a cada vizinho  $j$ , respeitando o raio de corte para as interações intermoleculares, se estiverem próximas o suficiente para considerar a interação, serão computadas as forças resultantes sobre cada uma. As novas posições e velocidades são calculadas utilizando algoritmos integradores das equações de movimento, o mais usual é o Verlet. O intervalo de armazenamento de trajetória não é igual ao passo de integração da simulação: para evitar desperdício da capacidade de armazenamento, a trajetória é gravada somente a cada centena ou mesmo milhar de passos de integração. Cada passo de integração da simulação determina um intervalo de tempo, e conjunto de passos forma a trajetória com as propriedades de interesse (FRENKEL; SMIT; RATNER, 1996; MARTÍNEZ; BORIN; SKAF, 2007).

### **3.3.2.1 Campos de Força**

A escolha adequada do campo de força é uma das prioridades para uma simulação de DM bem-sucedida devido a relação direta no cálculo dos potenciais e forças de interação inter e intramolecular que determinarão as trajetórias ao longo do tempo. Assim, Campo de Força é o conjunto de todos os parâmetros propostos para descreverem as interações intra e intermoleculares de um dado sistema. Diversos campos de força foram desenvolvidos e bem consolidados por grupos de pesquisa com aplicabilidades variadas a depender da complexidade do sistema: destacam-se o AMBER (CORNELL et al., 1995), CHARMM (BROOKS et al., 1983), GROMOS (VAN GUNSTEREN et al., 1996) e OPLS (JORGENSEN; TIRADO-RIVES, 1988) que apresentam boa parametrização para biomoléculas. Em geral, os átomos são considerados como pontos de massa que se movem podendo ser descritos

explicitamente (*all-atom*) ou simplificados por agrupamento (*united atom*), onde por exemplo átomos de hidrogênio ligados a carbono podem ser considerados um único sítio da molécula (CH), também chamado de *merged*.

A definição do campo de força baseia-se em um conjunto de equações chamadas de funções potenciais, para as quais são estabelecidos com parâmetros específicos para potenciais ligados relacionados as interações intramoleculares, e potenciais não-ligados para interações intermoleculares. As interações de longo alcance (eletrostáticas) decaem lentamente com a distância, podendo ser descritas por uma força efetiva obtida por métodos de soma infinita, onde o mais usual é a soma de Ewald no método SPME (*smooth particle mesh Ewald*) (MARTÍNEZ; BORIN; SKAF, 2007), sendo efetivo em precisão e velocidade de processamento.

### 3.3.2.2 Condições Periódicas de Contorno e Ensemble

O sistema definido para a simulação é delimitado por uma caixa de formato geométrico pré-estabelecido, e para que se aproxime mais da realidade são geralmente empregadas condições periódicas de contorno para eliminação das fronteiras através da replicação do sistema nas três dimensões. A caixa é então essencialmente replicada inúmeras vezes para criar uma rede periódica que tende ao limite termodinâmico, com número de moléculas (N) e volume (V) tendendo ao infinito, mas a razão N/V constante. De forma a minimizar o custo computacional e ter melhor controle das fronteiras é utilizada a convenção da imagem mínima, onde é considerada a interação apenas de curto alcance, dentro do limite de um raio de corte, a partir do qual potenciais que tendem a zero.

Um ensemble é um conjunto. Um ensemble canônico, por exemplo, é um conjunto de vários sistemas similares que, por estarem em contato diatérmico, compartilham a mesma temperatura, além de compartilharem também o número de partículas e volume (NVT). No contexto da simulação, usar um ensemble significa mimetizar este conjunto de sistemas vinculado a estas dadas restrições mediante o uso de parâmetros externos de controle (caixa rígida ou termostato ou barostato). Os parâmetros, em geral, são número de partículas (N), volume (V), pressão (P), temperatura (T) e energia total do sistema (E) que são mantidos constantes ou não a depender do sistema estudado.

No caso de sistemas biológicos, o mais usado é o ensemble isobárico-isotérmico (NPT), sendo N, T e P constantes ao longo da simulação e controlados por acoplamento de termostato e barostato enquanto o volume do sistema flutua (BERENDSEN et al., 1984). Os acoplamentos são definidos por métodos de escalonamento das velocidades ou dimensões da caixa, evitando flutuações dos parâmetros estabelecidos como constantes.

## 4 METODOLOGIA

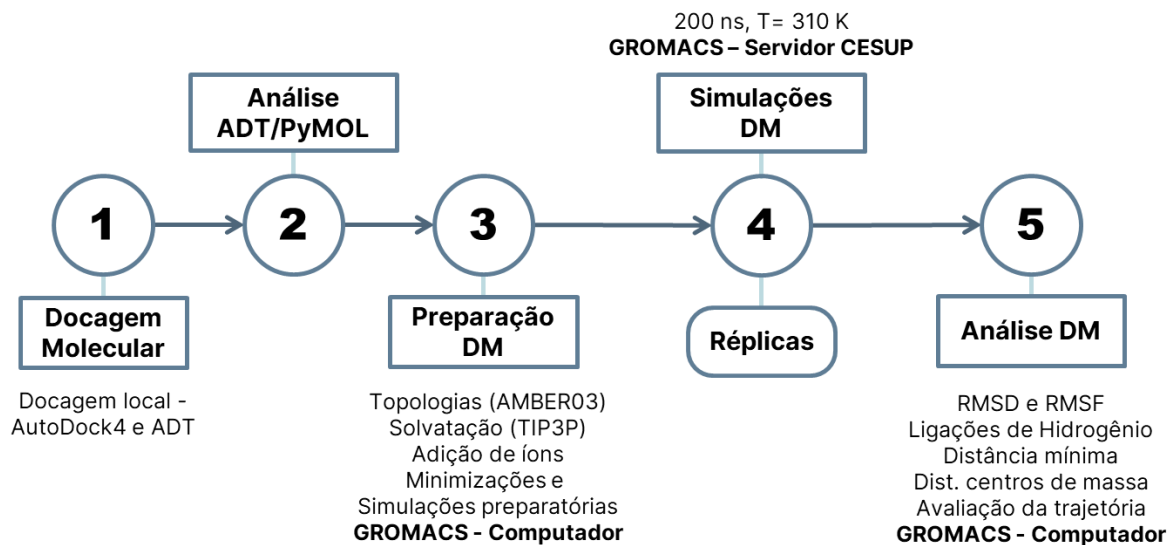
O presente estudo da interação do molnupiravir (EIDD-1931 5'-triphosphate) com as proteínas alvos RdRp (NSP12) e M<sup>PRO</sup> (NSP5) do SARS-CoV-2 foi realizado através de metodologias computacionais de Docagem e DM descritas no capítulo 3. A metodologia empregada (Figura 5) compreendeu a obtenção das estruturas por banco de dados, cálculos de Docagem seguidas de simulações de Dinâmica Molecular e análise dos resultados em cada etapa, sendo parte realizada no Laboratório do Grupo Química Teórica da UFRGS (GQT) e parte no CESUP (Centro Nacional De Super Computação).

### 4.1 BANCO DE DADOS

A molécula do molnupiravir (EIDD-1931 5'-triphosphate) foi obtida no banco de dados Pubchem no formato sdf e convertida para o formato pdb usando software OpenBabelGUI (O'BOYLE et al., 2011). O servidor online *Automatic Topology Builder* (ATB) (MALDE et al., 2011) foi utilizado para a minimização de energia do ligante usando o método B3LYP da teoria funcional da densidade (DFT) com base 6-31G\*, após a otimização da geometria, a estrutura do ligante 3D minimizada foi utilizada nas simulações de Docagem e DM.

As proteínas M<sup>PRO</sup> e RdRp do SARS-CoV-2 foram obtidas pelo banco de dados PDB (Protein Data Bank) gerenciada pela RCSB (Research Collaboratory for Structural Bioinformatics) com os identificadores 6lu7 e 6m71, respectivamente.

Figura 5. Esquema da metodologia geral utilizada em ordem de passos junto de breves descrições de pontos presentes.



## 4.2 DOCAGEM MOLECULAR

A preparação dos receptores e ligantes foi realizada na interface AutoDockTools-1.5.6 (MORRIS et al., 2009) e a ancoragem molecular do ligante MTP com os potenciais alvos (proteínas) do SARS-CoV-2, foi performada pelo programa AutoDock4 (MORRIS et al., 2009).

Antes do encaixe molecular, os átomos de hidrogênio polares de todas as enzimas e MTP foram retidos nas estruturas 3D, enquanto os hidrogênios apolares foram mesclados (do inglês, *merged*) e, em seguida, as cargas de Gasteiger foram atribuídas ao MTP (Morris et al., 2009; Sanner, 1999). Por outro lado, cargas de Kollmann foram atribuídas aos receptores. Na etapa de preparação, os termos torcionais (diedrais) de MTP foram deixados girar livremente e, finalmente, as estruturas otimizadas das proteínas alvo (rígidas) e do ligante MTP (flexível) foram salvos no formato PDBQT. O tamanho da caixa de grade (gridbox) ajustado antes do encaixe molecular foi definido usando coordenadas apropriadas para garantir que o MTP se encaixasse facilmente no sítio de interação das proteínas estudadas. As grades foram ajustadas com proporção de pontos de 52x44x58 para a protease M<sup>PRO</sup> e 45x55x60 para a proteína RdRp, com resolução de 0,375 Å. Esses tamanhos de caixa de grade definidos foram escolhidos para incluir todos os resíduos do sítio ativo de todas as enzimas estudadas (HU et al., 2022; XU et al., 2003). Em todas as

simulações de Docagem Molecular aplicadas foram definidos os parâmetros: algoritmo genético (100 execuções), número máximo de avaliações de energia (2.500.000) e número máximo de gerações (27.000) são selecionados. Valores padrão de 0,02 e 0,8 foram definidos para as taxas de mutação e crossover.

Após 100 execuções de ancoragem independentes, todos os modos de ligação encontrados de MTP a cada enzima foram agrupados pelo programa e classificados com base na energia livre de ligação mais negativa (Kcal/mol). Em seguida, as conformações do receptor "hit" obtidas, foram visualizadas qualitativamente e analisadas usando os programas AutoDockTools (ADT) e PyMOL (DELANO, 2002), que forneceram informações relevantes sobre o modo de ligação com a interface de ligação da proteína.

#### 4.3 DINÂMICA MOLECULAR

A DM clássica trata de simulações nas quais as forças entre as partículas são tratadas de modo clássico e não quântico. Essa metodologia é uma boa aproximação para o sistema que trata de interações entre receptor e ligante em solução com foco nas mudanças conformacionais. As simulações de DM neste trabalho foram todas realizadas com o pacote GROMACS (LINDAHL et al., 2022) e as estruturas iniciais obtidas a partir das poses de maior pontuação obtidas nas simulações de Docagem.

As coordenadas das proteínas (receptores) foram convertidas via Babel (O'BOYLE et al., 2011) para o formato pdb, e preparadas para as simulações de DM pela ferramenta pdb2gmx usando o pacote do GROMACS (ABRAHAM et al., 2015), empregando o campo de força AMBER03 (DUAN et al., 2003). Já para o ligante MTP, suas coordenadas foram primeiro preparadas com AutoDockTools (MORRIS et al., 2009) com a adição de átomos de hidrogênio, seguida pela conversão para formato pdb. Essas coordenadas foram submetidas ao servidor online Acypype (SOUSA DA SILVA; VRANKEN, 2012) para gerar arquivos de topologia GROMACS, escolhendo o General Amber Force Field, GAFF (WANG et al., 2004), e cargas parciais AM1-BCC (JAKALIAN; JACK; BAYLY, 2002), sendo o MTP completamente desprotonado com carga -4.

As coordenadas e topologias do molnupiravir e dos receptores foram novamente combinadas para construir as coordenadas e topologias dos complexos. A preparação do complexo para a simulação se inicia com a criação da caixa cubica, com faces a

1,2 nm de distância da proteína, e condições de contorno periódicas. A solvatação do sistema foi feita preenchendo a caixa com moléculas de água do tipo TIP3P (JORGENSEN et al., 1983; JORGENSEN; JENSON, 1998), onde foram necessárias cerca de 29.310 moléculas para o sistema M<sup>PRO</sup>+MTP e 94.944 moléculas de água para o sistema RdRp+MTP, e seguida pela inserção de íons de sódio e cloreto correspondentes à concentração fisiológica.

Esses sistemas foram inicialmente minimizados energeticamente usando método *Steepest Descent* (LINDAHL et al., 2019) e, em seguida, submetidos a uma simulação de 500 ps (250 mil passos de 2 fs cada) de duração com restrições nas posições do receptor e do ligante, permitindo que as moléculas de solvente e os íons equilibrem e relaxem no sistema. As interações de van der Waals foram calculadas usando um raio de corte de 1,1 nm e as interações eletrostáticas foram calculadas usando o método SPME (DARDEN; YORK; PEDERSEN, 1993; ESSMANN et al., 1995a, 1995b). Após isso, o sistema foi submetido, no CESUP, a uma sequência de três simulações de dinâmica molecular de termalização com 5 ns (2,5 milhões de passos de 2 fs cada) cada nas temperaturas de 200 K, 240 K e 280 K para simular a aproximação gradual e relaxamento do sistema todo sem restrições.

Concluídas essas etapas, foram realizadas simulações de produção com longa duração de 200 ns, no ensemble NPT, empregando um termostato V-rescale, semelhante ao acoplamento de Berendsen (BERENDSEN et al., 1984; BUSSI; DONADIO; PARRINELLO, 2007), e um barostato Parrinello-Rahman (PARRINELLO; RAHMAN, 1981). O passo de integração foi mantido 2 fs para cada um dos sistemas, totalizando 100 milhões de passos, o que em uma simulação direta levaria dias de processamento computacional mesmo no CESUP (até 7 dias para RdRp+MTP). Dessa forma, para melhorar gerenciamento, o sistema teve a trajetória dividida em 4 partes iguais. Os parâmetros para cada simulação foram mantidos comuns à todas as simulações, sendo as longas com raio de corte de 1,1 nm para interações do tipo Lennard-Jones e do tipo Coulomb, correção das interações eletrostáticas de longo alcance pelo método PME, temperatura de 310K mantida com o termostato V-rescale e pressão constante de 1,0133 bar (1 atm) mantida com o barostato de Parrinello-Rahman também.

As trajetórias, uma vez finalizadas, são obtidas e acopladas em um único arquivo preservando as estruturas do complexo apenas (proteína e ligante) de forma a ter um arquivo mais leve de processamento em computador local para as análises.

## 4.4 ANÁLISE DOS RESULTADOS

Os resultados de docagem foram obtidos a partir da análise de todos os clusters gerados pelo AutoDock, realizadas com o auxílio de laboratório parceiro da Turquia, onde foram selecionados e cedidos os resultados de menor energia com a pose resultante para sequência do projeto.

As trajetórias obtidas somando 200 ns de simulação foram convertidos em um arquivo de trajetória único, foram realizadas as análises com os módulos do pacote GROMACS padrão, como a evolução da estrutura média por RMSD, mobilidade relativa de segmentos RMSF, análise de ligações de hidrogênio intermolecular, distâncias entre os centros de massa do ligante e da proteína junto com a distância mínima entre eles, e acompanhado dos dados análise visual e qualitativa das trajetórias realizada com o VMD (HUMPHREY; DALKE; SCHULTEN, 1996). Os gráficos e figuras contidos neste trabalho foram gerados pelos programas VMD, PyMol (DELANO, 2002) e xmgrace (TURNER, 2005).

## 5 RESULTADOS E DISCUSSÕES

### 5.1 DOCAGEM MOLECULAR

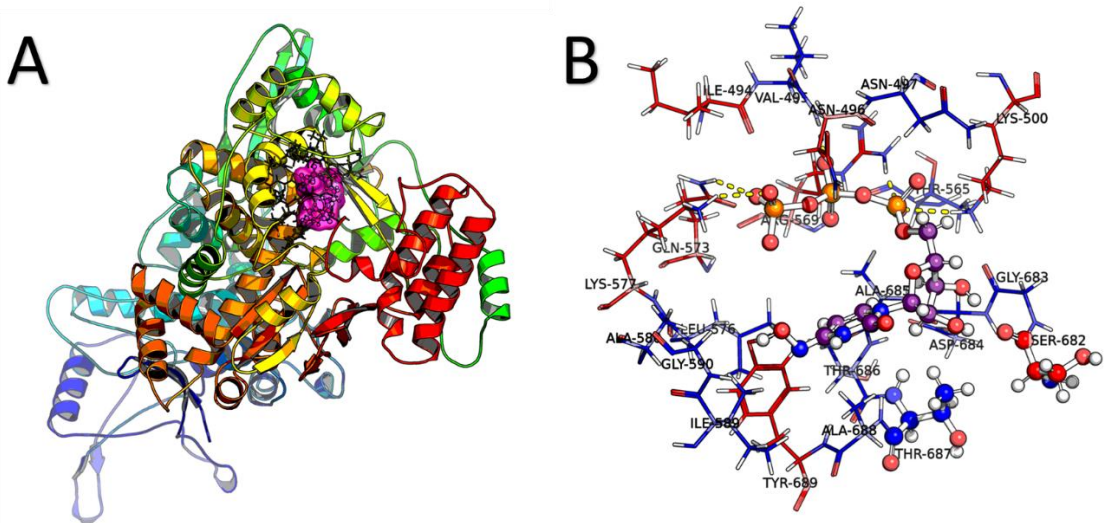
O estudo de encaixe ligante-macromolécula por Docagem Molecular é uma técnica chave usada na biologia molecular com o objetivo de prever os prováveis modos de ligação para um ligante que complexa com uma macromolécula de estrutura tridimensional conhecida. De acordo com a metodologia citada no capítulo anterior, foram realizadas docagens do análogo de ribonucleosídeo  $\beta$ -D-N4-hidroxicitidina (NHC) na forma ativa 5'-trifosfato (MTP) com seu principal alvo RNA polimerase dependente de RNA (RdRp) e com a Protease Principal ( $M^{PRO}$ ). Os resultados da docagem das conformações enzima-ligante são dispostos em clusters de acordo com a similaridade geométrica (categorizados para um desvio médio quadrático padrão máximo de 2,0 Å) e energética: dessa forma o cluster que representa o de menor energia é considerado o mais favorável.

O estudo de docagem selecionou os resultados de menor energia compostos pela população do primeiro cluster. Para o sistema MTP com RdRp, foi obtido a



energia livre de ligação de  $-7,86$  Kcal/mol, com uma população de 5 no primeiro cluster e constante de inibição  $1,74$   $\mu$ M. Esse complexo foi analisado no PyMOL, onde foram encontradas 21 interações, sendo pelo menos 8 contatos polares computados com os resíduos ARG-569, ASN-496, GLN-573, ILE-494, LYS-500, LYS-577, **SER-682** e TYR-689, e 13 contatos não polares ou hidrofóbicos com ALA-580, ALA-685, ALA-688, ASN-497, ASP-684, GLY-590, GLY-683, ILE-589, LEU-576, THR-565, THR-686, **THR-687** e VAL-495. Dentre esses contatos os resíduos SER-682 e THR-687 compõem os principais resíduos de reconhecimento e seleção do trifosfato de nucleosídeo (NTP) na replicação viral (Figura 6).

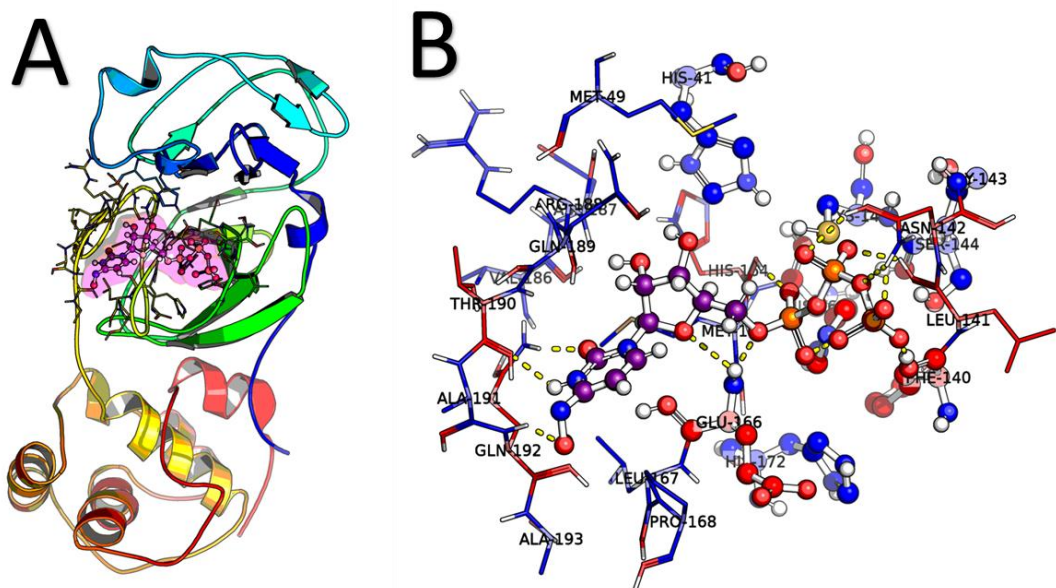
Figura 6 . Visão tridimensional do complexo formado pelo ligante MTP e a RdRp através de estudos de Docagem Molecular, sendo (A) a proteína representada em estruturas secundárias, resíduos explícitos no sítio de interação com o MTP representado por superfície roxa. (B) Visão aproximada do sítio ligante com os resíduos de interação explícitos e coloridos de vermelho (contato polar) e azul (não polar) e em destaque com cadeia maior o MTP e resíduos principais.



Já o sistema composto pela  $M^{PRO}$  e o ligante MTP gerou o primeiro cluster com energia livre de ligação igual à  $-6,19$  Kcal/mol em uma população de 16 conformações e constante de inibição de  $28,81$   $\mu$ M. O complexo de  $M^{PRO}$  analisado no PyMOL apresentou um número maior que na RdRp de possíveis interações na cavidade da macromolécula com 23 interações, sendo 8 contatos polares com os resíduos ASN-142, GLN-192, **GLU-166**, **HIS-163**, HIS-164, LEU-141, **PHE-140** e THR-190, e também 15 contatos não polares ou hidrofóbicos com os resíduos ALA-191, ALA-193,

ARG-188, ASP-187, **CYS-145**, GLN-189, **GLY-143**, **HIS-172**, **HIS-41**, LEU-167, MET-165, MET-49, PRO-168, **SER-144** e VAL-186. Sendo destacados todos os resíduos definidos como essenciais, nas sub cavidades S1 e S2 da M<sup>PRO</sup> (Figura 7).

Figura 7 . Visão tridimensional do complexo formado pelo ligante MTP e a MPRO através de estudos de Docagem Molecular, sendo (A) a proteína representada em estruturas secundárias, resíduos explícitos no sítio de interação com o MTP representado por superfície roxa. (B) Visão aproximada do sítio ligante com os resíduos de interação explícitos e coloridos de vermelho (contato polar) e azul (não polar) e em destaque com cadeia maior o MTP e resíduos principais.



Como os resultados de Docagem Molecular não consideraram a flexibilidade da macromolécula frente ao ligante, o objetivo dessa técnica consistiu em obter os complexos de interação favoráveis para os estudos complementares por DM.

## 5.2 DINÂMICA MOLECULAR

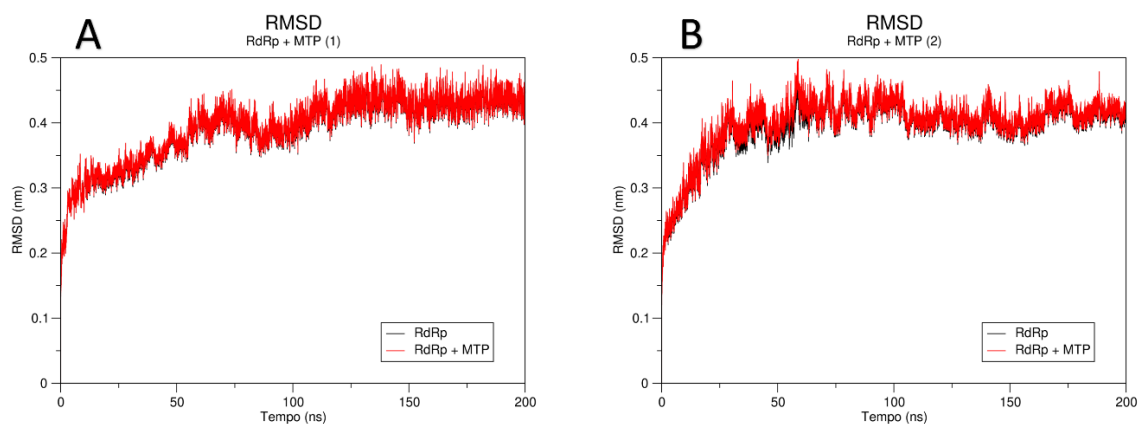
O estudo dos sistemas por DM visou levantar dados em torno da permanência ou não do complexo de interação resultante da docagem realizada anteriormente. Dessa forma, os sistemas foram tratados como descritos na metodologia e as simulações de 200ns feitas em duplicatas para melhor comparação de resultados. A duplicata é gerada pela mudança da “seed” e habilitação de geração de velocidades nos primeiros 50ns da nova trajetória replicada.

Os arquivos de trajetórias para cada 50ns de simulação foram agrupados em um arquivo único, corrigindo a trajetória em relação a posição da proteína na caixa de simulação de forma a conseguir avaliar visualmente o comportamento do sistema ao longo da simulação e extrair dados relevantes para caracterização do modo de ligação no complexo. As análises realizadas foram de RMSD da proteína e do complexo em relação a posição inicial (0ns), RMSF referenciando o C<sub>alfa</sub> da cadeia durante o todo da simulação e parcial (região equilibrada), número de ligações de hidrogênio e distâncias entre os centros de massa do ligante e da proteína junto com a distância mínima entre eles também.

### 5.2.1 RNA polimerase dependente de RNA (RdRp) e MTP

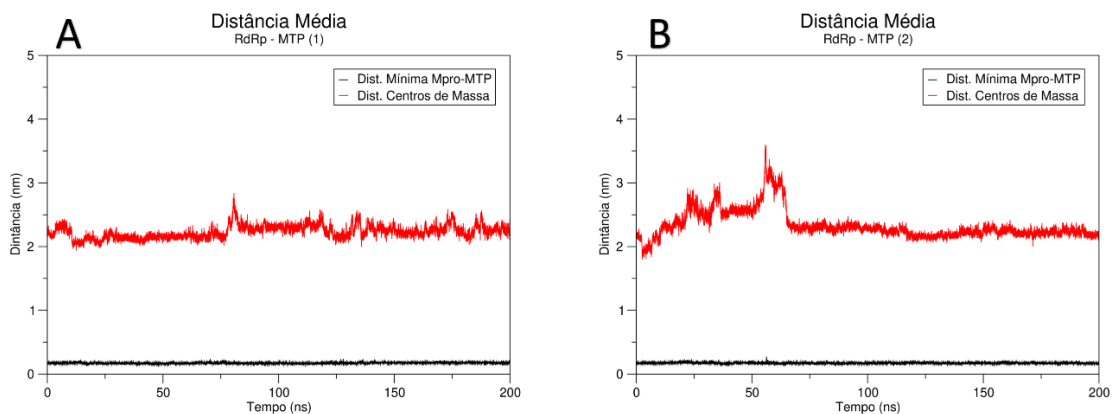
Os resultados de RMSD obtidos pela trajetória da primeira replicata reproduzida pelo pacote GROMACS do sistema RdRp e MTP (Figura 8, A) demonstraram que o sistema possuiu flutuações dentro da caixa solvatada de simulação até 100ns e após entra em equilíbrio até o final, sendo esse dado corroborado pela segunda replicata (Figura 8, B) que apresentou maior flutuação inicial, mas atinge o equilíbrio da mesma forma. Observa-se na comparação dos RMSDs calculados em relação ao complexo e a proteína RdRp livre devem permanecer alinhados para confirmar que não houve fuga do ligante ou problemas durante a simulação.

Figura 8 . Resultados gráficos do cálculo de RMSD para o sistema RdRp e MTP, sendo (A) primeira replicata e (B) segunda replicata.



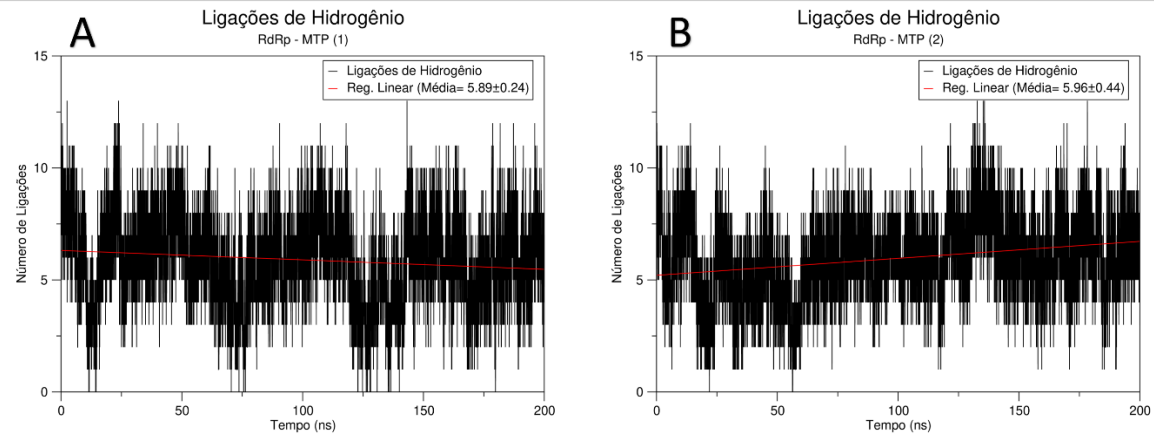
As distâncias mínimas entre o MTP e a RdRp (Figura 9) para as duas réplicas apresentaram estabilidade ao longo de toda a simulação, comprovando com a visualização da trajetória que o ligante permanece próximo a proteína. Porém, nos resultados da distância entre os centros de massa da proteína e do ligante os resultados divergiram: enquanto que para a primeira réplica houve maior permanência com desvios em 75 e 125ns indicando uma mudança da conformação, mas se mantendo próximo do sítio de interação inicial, a segunda réplica mostra uma grande variância de conformações logo no início da simulação causando a migração do MTP em torno da proteína com afastamento dos centros de massa em até 1,5 Å, seguida de estabilização do complexo até o fim da simulação, confirmado visualmente no VMD.

Figura 9. Resultados gráficos do cálculo de distância mínima e entre os centros de massa para o sistema RdRp e MTP, sendo (A) primeira replicata e (B) segunda replicata.



Ao longo dos 200ns de simulação do sistema, com liberdade de movimentos, interações e ligações de hidrogênio puderam ser formadas e desfeitas. A análise do número de ligações de hidrogênio em função do tempo é um indicativo poderoso da tendência de modos de ligação com sítios polares ou hidrofóbicos. O gráfico obtido (Figura 10) para a primeira réplica demonstra uma regressão linear com coeficiente angular negativo e média próxima de 6, enquanto que a segunda réplica, mesmo apresentando um coeficiente angular positivo, com média próxima de 6. Apesar das pequenas discordâncias entre elas, a média flutua próxima de 6 ligações de hidrogênio possíveis entre o ligante e a proteína.

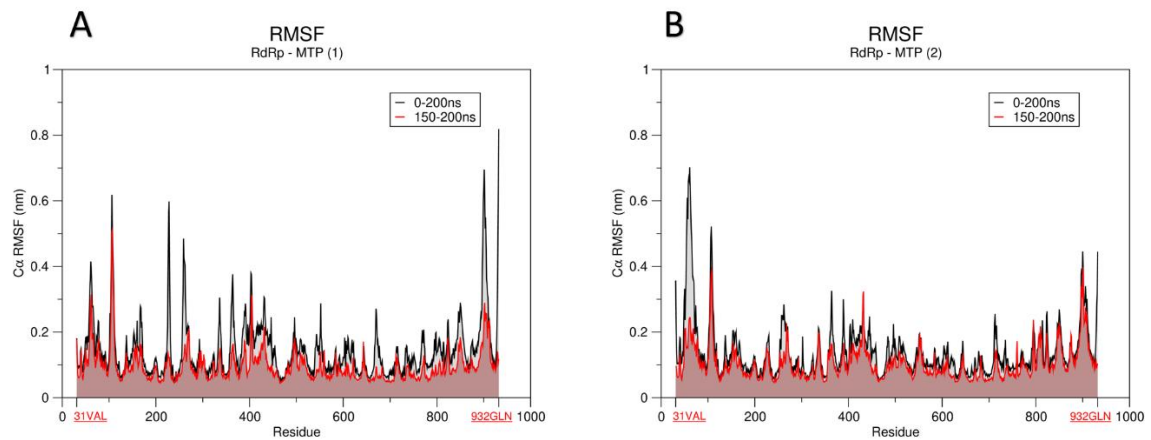
Figura 10. Resultados gráficos das ligações de hidrogênio para o sistema RdRp e MTP, sendo (A) primeira replicata e (B) segunda replicata.



As análises de RMSF determinam a flutuação de um átomo ou grupo de átomos durante a simulação, fornecendo um parâmetro de flexibilidade dos resíduos dentro da proteína e por esse motivo o cálculo é feito em referência ao esqueleto, composto pelos carbonos alfa, da proteína com seus resíduos. Uma região caracterizada por altos valores de RMSF indica alta divergência da posição média e, portanto, alta mobilidade estrutural. Para o sistema RdRp e MTP foram realizados dois cálculos de RMSF, um durante o período de equilíbrio de interação ligante e proteína definido pelo RMSD (de 150 até 200ns) e outro ao longo de toda a simulação.

Os resultados de RMSF (Figura 11) para a primeira replicata do sistema demonstrou uma semelhança nos resíduos que se mantêm flexíveis nos dois tempos calculados havendo uma diminuição da mobilidade no tempo de equilíbrio. A segunda réplica acompanha a tendência da primeira, com a diferença de uma grande diminuição da mobilidade perto da terminação N' em torno dos resíduos 31-50.

Figura 11. Resultados gráficos do cálculo de RMSF para o sistema RdRp e MTP, sendo (A) primeira replicata e (B) segunda replicata.



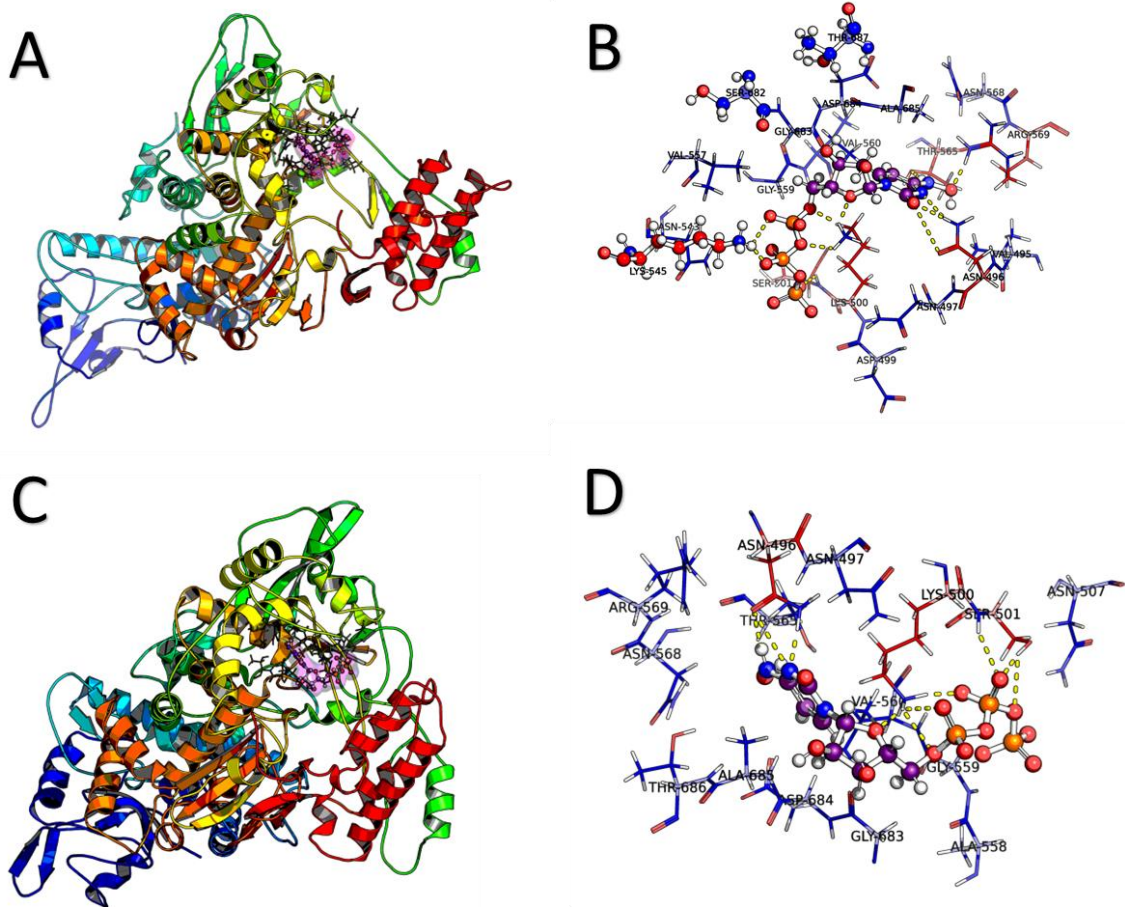
Analisando os gráficos dos resultados e trajetórias obtidas nas replicatas, foram definidos dois momentos da simulação de cada réplica para investigação de possíveis interações do MTP com resíduos na vizinhança e sua comparação com os resíduos apontados pela simulação de docagem (Figura 12). No caso da RdRp com MTP, foram investigados os tempos de 100 e 200ns para elucidar o que mudou entre os 100 primeiros nanossegundos e o final da simulação. Os resíduos destacados em negrito são relatados como responsáveis pelo reconhecimento do análogo nucleosídeo para incorporação na replicação do RNA viral (WU et al., 2020).

Para a primeira replicata, no tempo de 100ns da simulação, verificou-se um total de 19 interações em torno do MTP, sendo 6 polares com os resíduos ARG-569, ASN-496, LYS-500, **LYS-545**, SER-501 e THR-565, e outras 13 interações não polares com os resíduos ALA-685, ASN-497, ASN-543, ASN-568, ASP-499, ASP-684, GLY-559, GLY-683, **SER-682**, **THR-687**, VAL-495, VAL-557, e VAL-560. Para a segunda réplica, no tempo de 100ns da simulação, observou-se um total de 15 interações, sendo apenas 3 delas contatos polares com os resíduos ASN-496, LYS-500 e SER-501, enquanto que as outras 12 não polares ocorrem com os resíduos ALA-558, ALA-685, ARG-569, ASN-497, ASN-507, ASN-568, ASP-684, GLY-559, GLY-683, THR-565, THR-686, VAL-560. Na primeira réplica foram mantidos 11 resíduos de interação da posição inicial de docagem, outros 8 foram incluídos, dentre eles o resíduo **LYS-545** que é responsável pela ligação e posicionamento do NTP para replicação. Enquanto que na segunda replicata apenas 9 resíduos de interação foram



conservados e mesmo estando próximos do sítio catalítico não foram computadas interações entre os principais resíduos envolvidos na replicação.

Figura 12. Visão tridimensional do complexo formado pelo ligante MTP e a RdRp através de estudos de DM no tempo de 100ns, sendo (A) primeira replicata e a (C) segunda replicata com a proteína representada em estruturas secundárias, resíduos no sítio de interação com o MTP representado por superfície roxa. (B) primeira replicata e a (D) segunda replicata com visão aproximada do sítio ligante com os resíduos de interação explícitos e coloridos de vermelho (contato polar) e azul (não polar) e em destaque com cadeia maior o MTP e resíduos principais.



No tempo de 200ns de simulação (Figura 13), a primeira replicata apresentou permanência de 8 resíduos de interação em relação a docagem, mantendo um total 15 interações ao final da simulação, sendo 6 polares com os resíduos ALA-558, ARG-569, GLY-559 LYS-500, **LYS-545** e **SER-682**, e também 9 interações não polares com os resíduos ALA-685, ASN-496, ASN-543, ASP-684, GLN-573, GLY-683, **THR-687**, VAL-557 e VAL-560. Já a segunda replicata, mantendo 8 interações iniciais, apresentou um aumento de 15 para 17 interações no sítio da enzima, tendo 5

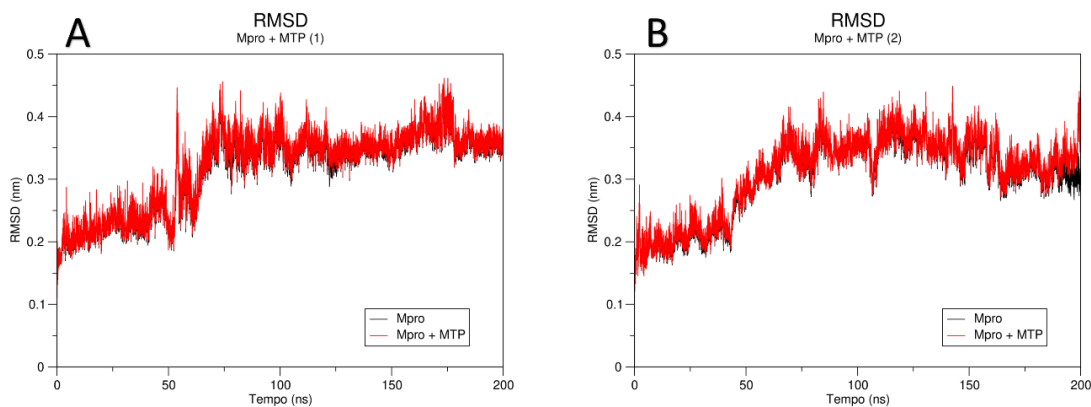




### 5.2.2 Protease Principal (M<sup>PRO</sup>) e MTP

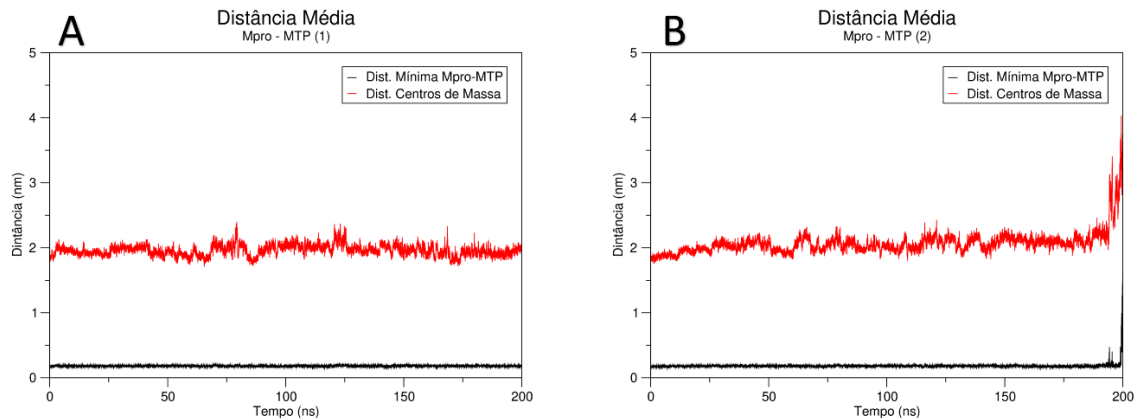
As simulações de trajetória de DM para o sistema M<sup>PRO</sup> e MTP em duplicata apresentaram gráficos de RMSD (Figura 14) com variação considerável em relação a posição inicial até 100ns, o que é explicável pelo grande movimento do complexo na caixa solvatada de simulação indicando que estes sistemas, mesmo com as etapas de termalização, não havia atingido o equilíbrio. Ao final do gráfico da segunda replicata (Figura 14, B) é possível observar uma divergência entre o RMSD da proteína e do complexo, indicando o afastamento do ligante da interface da proteína, o que não foi observado na primeira replicata (Figura 14, A) ou nos complexos de RdRp com MTP (Figura 8).

Figura 14. Resultados gráficos do cálculo de RMSD para o sistema MPRO e MTP, sendo (A) primeira replicata e (B) segunda replicata.



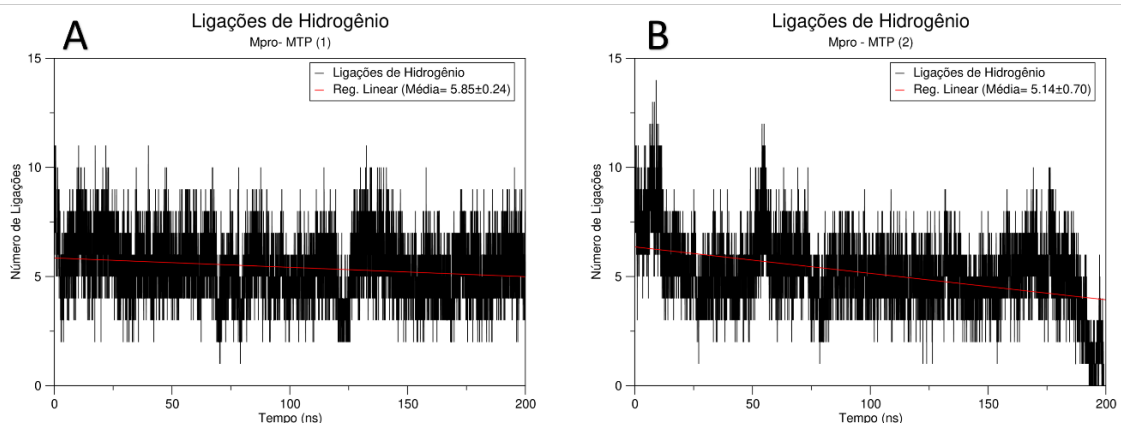
Os resultados do cálculo de distâncias dos centros de massa e entre a superfície do ligante com a proteína demonstraram estabilidade ao longo da simulação, com exceção da segunda replicata que novamente apresenta indícios de fuga ao se aproximar dos 200ns de simulação (Figura 15). A tendência de permanência do MTP nas imediações da proteína em grandes períodos de tempo da simulação indica a existência de uma interação favorável.

Figura 15. Resultados gráficos do cálculo de distância mínima e entre os centros de massa para o sistema MPRO e MTP, sendo (A) primeira replicata e (B) segunda replicata.



Os gráficos do número de ligações de hidrogênio pelo tempo (Figura 16) apresentaram, para as duas réplicas, coeficiente angular negativo e média próxima de 6. No entanto, para a segunda replicata o número de ligações de Hidrogênio têm um decréscimo muito maior (até próximo de zero) explicado pela discrepância dos pontos iniciais e finais. Caso fossem desconsiderados restaria uma faixa de simulação com comportamento próximo ao da primeira replicata (de 25 até 175ns).

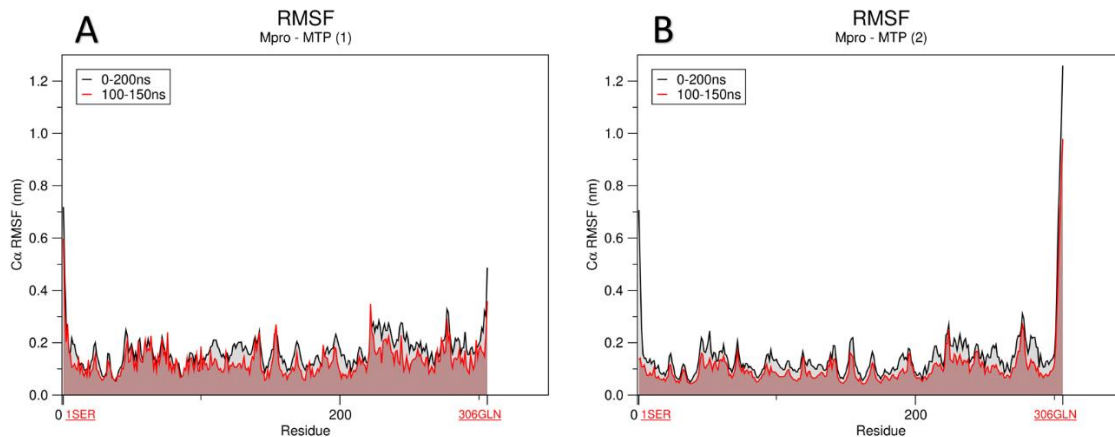
Figura 16. Resultados gráficos das ligações de hidrogênio para o sistema MPRO e MTP, sendo (A) primeira replicata e (B) segunda replicata.



Para as análises de flexibilidade dos resíduos em relação ao esqueleto composto dos carbonos alfa da proteína, as duas replicatas apresentaram pouca diferença na mobilidade dos resíduos no interior da proteína, porém grande mobilidade nos extremos. No caso da primeira replicata, o laço terminal N' com o resíduo SER-1

possui o máximo de mobilidade, enquanto que na segunda replicata a terminação C da cadeia enzimática demonstrou maior mobilidade variando em até 1,2 nm no resíduo GLN-306 (Figura 17).

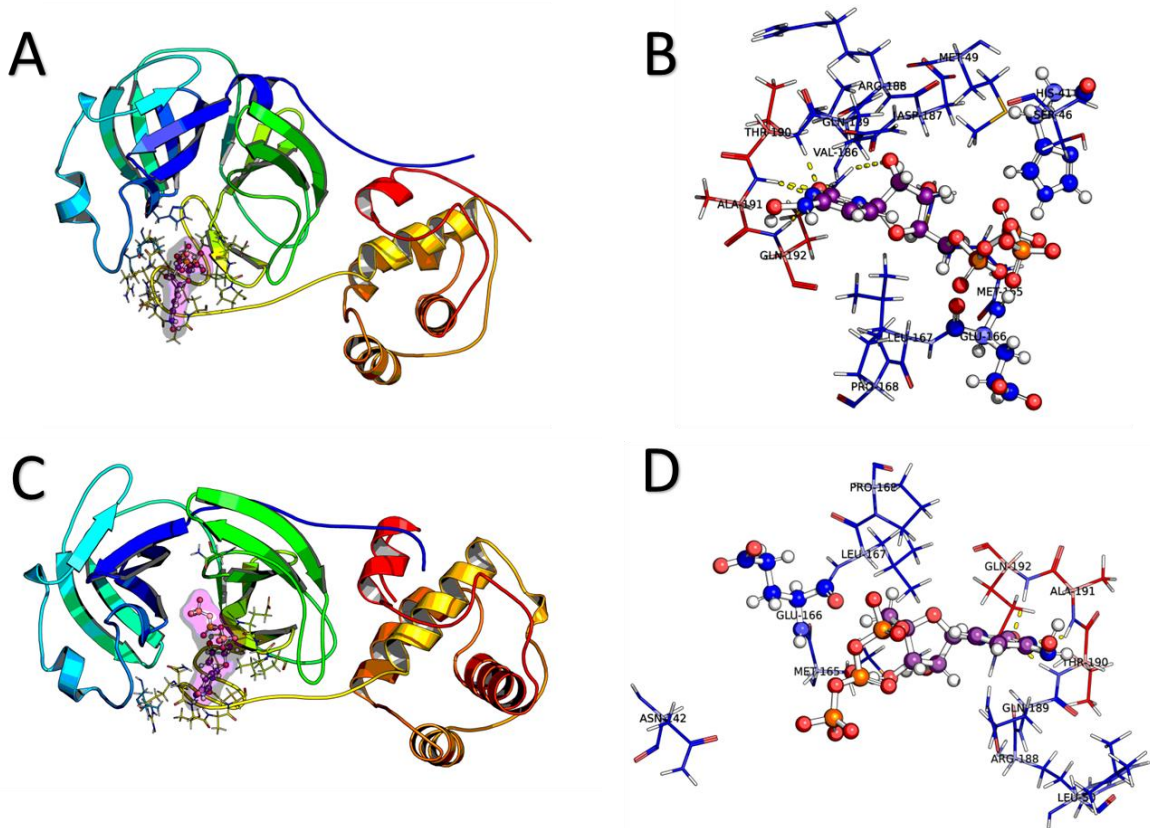
Figura 17. Resultados gráficos do cálculo de RMSF para o sistema MPRO e MTP, sendo (A) primeira replicata e (B) segunda replicata.



Os resíduos HIS-41 e CYS-145 são os mais comentados pela literatura (HU et al., 2022) por atuarem no sítio catalítico da enzima protease principal ( $M^{PRO}$ ) do SARS-CoV-2. Por esse motivo também foram investigados os resíduos de interação em dois momentos da simulação de DM, sendo definido no tempo 100 e 200ns para a primeira réplica e 100 e 175ns para a segunda réplica. O tempo de 175ns para a segunda réplica foi definido pela falta de um complexo no tempo de 200ns, quando ocorre divergência na simulação e o ligante escapa da superfície da proteína indicando abandono do sítio de interação, sendo esse dado confirmado visualmente com VMD.

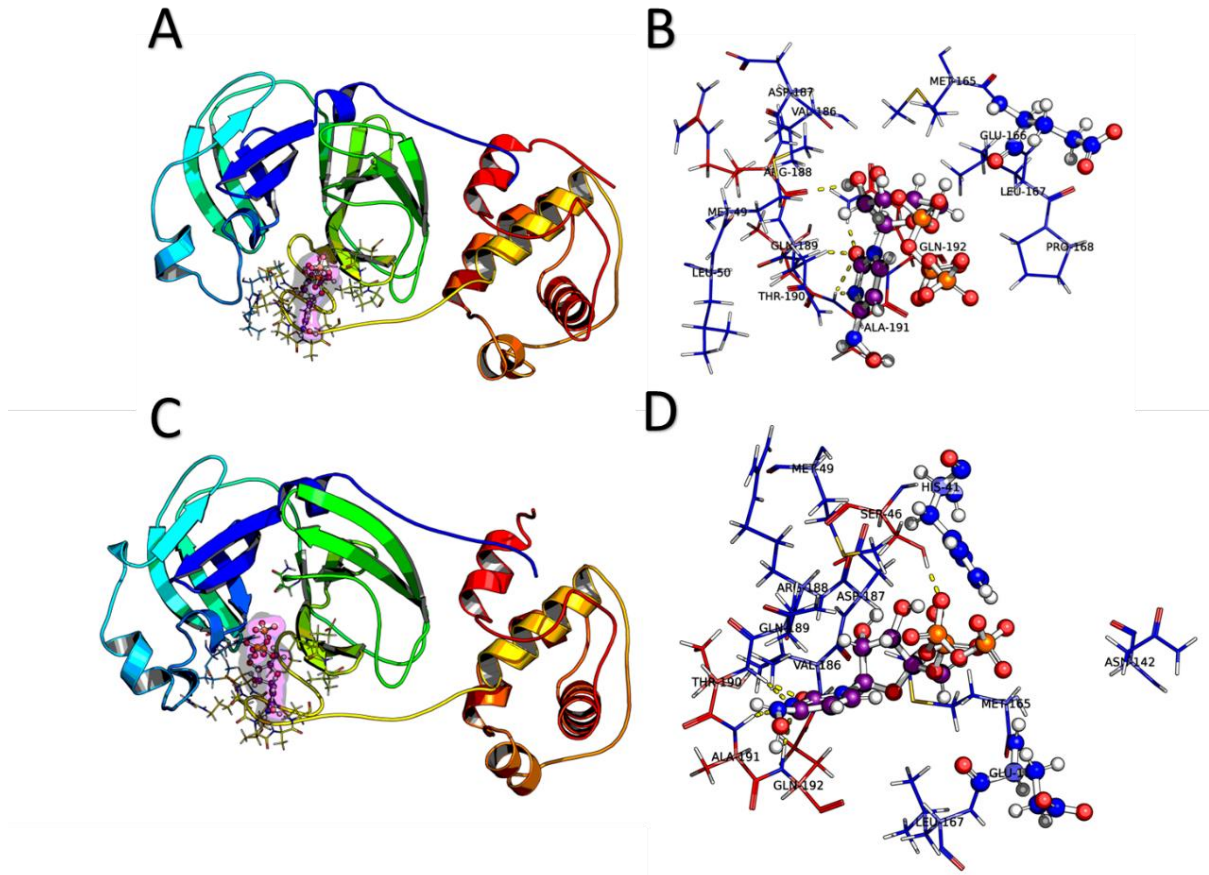
Na docagem o PyMOL computou cerca de 23 interações no total, enquanto que na primeira réplica em 100ns definiu 14 interações possíveis, sendo 3 polares com ALA-191, GLN-192 e THR-190, e também outras 11 não polares com **HIS-41**, ARG-188, ASP-187, GLN-189, **GLU-166**, LEU-167, MET-165, MET-49, PRO-168, SER-46 e VAL-186. Já na segunda réplica em 100ns existem 11 interações no total, sendo dessas as mesmas 3 polares que na primeira réplica e outras 8 interações não polares ARG-188, ASN-142, GLN-189, **GLU-166**, LEU-167, LEU-50, MET-165 e PRO-168 (Figura 18). Houve uma redução de resíduos que aparentam interagir com o MTP nas duas replicatas em relação à docagem, mantendo interações com até 2 resíduos principais do sítio de interação da proteína.

Figura 18. Visão tridimensional do complexo formado pelo ligante MTP e a MPRO através de estudos de DM no tempo de 100ns, sendo (A) primeira replicata e a (C) segunda replicata com a proteína representada em estruturas secundárias, resíduos no sítio de interação com o MTP representado por superfície roxa. (B) primeira replicata e a (D) segunda replicata com visão aproximada do sítio ligante com os resíduos de interação explícitos e coloridos de vermelho (contato polar) e azul (não polar) e em destaque com cadeia maior o MTP e resíduos principais.



Nos 200ns da primeira réplica do sistema  $M^{PRO}$  e MTP foram identificadas cerca de 13 interações, sendo 4 polares com os resíduos ALA-191, ARG-188, GLN-192, THR-190, e também outras 9 interações não polares com os resíduos ASP-187, GLN-189, **GLU-166**, LEU-167, LEU-50, MET-165, MET-49, PRO-168, VAL-186. Já para os 175ns na segunda réplica, foram computados 4 contatos polares com ALA-191, GLN-192, SER-46 e THR-190, e outros 10 não polares com os resíduos ARG-188, ASN-142, ASP-187, GLN-189, **GLU-166**, **HIS-41**, LEU-167, MET-165, MET-49 e VAL-186 (Figura 19). A convergência de resultados se mantém, comparando os tempos próximos do final da simulação, demonstrando frequência de interações com o resíduo **GLU-166**, um dos citados do sub sítio S1 da proteína.

Figura 19. Visão tridimensional do complexo formado pelo ligante MTP e a MPRO através de estudos de DM no tempo de 200ns para a primeira replicata e 175ns na segunda, sendo (A) primeira replicata e a (C) segunda replicata com a proteína representada em estruturas secundárias, resíduos no sítio de interação com o MTP representado por superfície roxa. (B) primeira replicata e a (D) segunda replicata com visão aproximada do sítio ligante com os resíduos de interação explícitos e coloridos de vermelho (contato polar) e azul (não polar) e em destaque com cadeia maior o MTP e resíduos principais.



## 6 CONSIDERAÇÕES FINAIS

No presente trabalho foi aprofundado o estudo de prováveis interações do pró fármaco Molnupiravir em sua forma ativa 5'-trifosfato, recentemente aprovado para uso emergencial pela ANVISA no Brasil para Covid19, com a RNA polimerase dependente de RNA (RdRp) e a Protease Principal ( $M^{PRO}$ ) para inibição da replicação do vírus SARS-CoV-2.

Ambos os resultados de Docagem Molecular demonstraram grande afinidade através da energia livre de ligação bastante favorável com RdRp de -7.86 Kcal/mol, enquanto que a  $M^{PRO}$  apresentou -6.19 Kcal/mol.

Nas simulações de Dinâmica Molecular (DM) foram determinadas interações polares significativas com resíduos no sítio ativo da RdRp, que são responsáveis pelo reconhecimento da estrutura de trifosfato de nucleosídeo (NTP) viral para replicação. Bem como a permanência do ligante no sítio ativo, o que foi constatado como provável, uma vez que não houve fugas recorrentes do sítio ativo e os resíduos de interação são recorrentes entre as replicatas ao longo das simulações. Também foram determinadas interações favoráveis, mesmo que inferiores as computadas por docagem, com a  $M^{PRO}$  no sítio catalítico composto por HIS-41 e CYS-145, mas o resíduo participante mais recorrente foi o GLU-166. A possível fuga ao final da segunda replicata foi considerada como irregular em comparação com o conjunto de dados obtidos.

Dessa forma, o molnupiravir demonstrou grande afinidade com seu principal alvo RdRp e com um possível alvo em paralelo  $M^{PRO}$ .

## REFERÊNCIAS

- ABRAHAM, M. J. et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. **SoftwareX**, v. 1, p. 19–25, 2015.
- ALLOUCHE, A. Software News and Updates Gabedit — A Graphical User Interface for Computational Chemistry Softwares. **Journal of computational chemistry**, v. 32, p. 174–182, 2012.
- ANDERSEN, K. G. et al. The proximal origin of SARS-CoV-2. **Nature Medicine**, v. 26, n. 4, p. 450–452, 2020.
- Anvisa aprova venda do medicamento Lagevrio (molnupiravir) em farmácias — Agência Nacional de Vigilância Sanitária - Anvisa.** Disponível em: <<https://www.gov.br/anvisa/pt-br/assuntos/noticias-anvisa/2022/anvisa-aprova-venda-do-medicamento-lagevrio-molnupiravir-em-farmacias>>. Acesso em: 8 abr. 2023.
- AutoDock.** Disponível em: <<https://autodock.scripps.edu/>>. Acesso em: 8 abr. 2023.
- BERENDSEN, H. J. C. et al. Molecular dynamics with coupling to an external bath. **The Journal of chemical physics**, v. 81, n. 8, p. 3684–3690, 1984.
- BROOKS, B. R. et al. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. **Journal of computational chemistry**, v. 4, n. 2, p. 187–217, 1983.
- BUSSI, G.; DONADIO, D.; PARRINELLO, M. Canonical sampling through velocity rescaling. **The Journal of chemical physics**, v. 126, n. 1, p. 014101, 2007.
- CALLIGARI, P. et al. Molecular investigation of SARS–COV-2 proteins and their interactions with antiviral drugs. **Viruses**, v. 12, n. 4, 1 abr. 2020.
- CERAOLO, C.; GIORGI, F. M. Genomic variance of the 2019-nCoV coronavirus. **Journal of Medical Virology**, v. 92, n. 5, p. 522–528, 2020.
- 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.
- COVID - Coronavirus Statistics - Worldometer.** Disponível em: <<https://www.worldometers.info/coronavirus/>>. Acesso em: 8 abr. 2023.



- DARDEN, T.; YORK, D.; PEDERSEN, L. Particle mesh Ewald: An  $N \cdot \log(N)$  method for Ewald sums in large systems. **The Journal of chemical physics**, v. 98, n. 12, p. 10089–10092, 1993.
- DELANO, W. L. Pymol: An open-source molecular graphics tool. **CCP4 Newsl. Protein Crystallogr**, v. 40, n. 1, p. 82–92, 2002.
- DUAN, Y. et al. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. **Journal of computational chemistry**, v. 24, n. 16, p. 1999–2012, 2003.
- ESSMANN, U. et al. A smooth particle mesh Ewald method. **The Journal of chemical physics**, v. 103, n. 19, p. 8577–8593, 1995a.
- ESSMANN, U. et al. A smooth particle mesh Ewald method. **The Journal of chemical physics**, v. 103, n. 19, p. 8577–8593, 1995b.
- FRENKEL, D.; SMIT, B.; RATNER, M. A. **Understanding molecular simulation: from algorithms to applications**. [s.l.] Academic press San Diego, 1996. v. 2
- GUO, Z.-D. et al. Aerosol and surface distribution of severe acute respiratory syndrome coronavirus 2 in hospital wards, Wuhan, China, 2020. **Emerging infectious diseases**, v. 26, n. 7, p. 1586, 2020.
- HILGENFELD, R. **From SARS to MERS: crystallographic studies on coronaviral proteases enable antiviral drug design**. **The FEBS journal**, 1 set. 2014.
- HU, Q. et al. **The SARS-CoV-2 main protease (Mpro): Structure, function, and emerging therapies for COVID-19**. **MedComm** John Wiley and Sons Inc, , 1 set. 2022.
- HUMPHREY, W.; DALKE, A.; SCHULTEN, K. VMD: visual molecular dynamics. **Journal of molecular graphics**, v. 14, n. 1, p. 33–38, 1996.
- JAKALIAN, A.; JACK, D. B.; BAYLY, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. **Journal of computational chemistry**, v. 23, n. 16, p. 1623–1641, 2002.
- JORGENSEN, W. L. et al. Comparison of simple potential functions for simulating liquid water. **The Journal of chemical physics**, v. 79, n. 2, p. 926–935, 1983.
- JORGENSEN, W. L.; JENSON, C. Temperature dependence of TIP3P, SPC, and TIP4P water from NPT Monte Carlo simulations: Seeking temperatures of maximum density. **Journal of computational chemistry**, v. 19, n. 10, p. 1179–1186, 1998.
- JORGENSEN, W. L.; TIRADO-RIVES, J. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic



- peptides and crambin. **Journal of the American Chemical Society**, v. 110, n. 6, p. 1657–1666, 1988.
- KABINGER, F. et al. Mechanism of molnupiravir-induced SARS-CoV-2 mutagenesis. **Nature Structural and Molecular Biology**, v. 28, n. 9, p. 740–746, 1 set. 2021.
- KUSHWAHA, P. P. et al. Identification of Natural Inhibitors Against SARS-CoV-2 Drugable Targets Using Molecular Docking, Molecular Dynamics Simulation, and MM-PBSA Approach. **Frontiers in Cellular and Infection Microbiology**, v. 11, 12 ago. 2021.
- LEŚ, A.; ADAMOWICZ, L.; RODE, W. Structure and conformation of N4-hydroxycytosine and N4-hydroxy-5-fluorocytosine. A theoretical ab initio study. **Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression**, v. 1173, n. 1, p. 39–48, 1993.
- LINDAHL, E. et al. GROMACS 2019.4 Manual (Version 2019.4). **Zenodo**, 2019.
- LINDAHL, E. et al. GROMACS 2019 Source code (2018). **URL <https://doi.org/10.5281/zenodo>**, v. 2424363, 2022.
- LOMBARDINO, J. G.; LOWE, J. A. **The role of the medicinal chemist in drug discovery - Then and now. Nature Reviews Drug Discovery**, out. 2004.
- MALDE, A. K. et al. An automated force field topology builder (ATB) and repository: version 1.0. **Journal of chemical theory and computation**, v. 7, n. 12, p. 4026–4037, 2011.
- MALONE, B.; CAMPBELL, E. A. **Molnupiravir: coding for catastrophe. Nature Structural and Molecular Biology** Nature Research, , 1 set. 2021.
- MARTÍNEZ, L.; BORIN, I. A.; SKAF, M. S. Fundamentos de simulação por dinâmica molecular. **Métodos de Química Teórica e Modelagem Molecular**, p. 413–452, 2007.
- MORRIS, G. M. et al. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. **Journal of Computational Chemistry**, v. 19, n. 14, p. 1639–1662, 15 nov. 1998.
- MORRIS, G. M. et al. Software news and updates AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. **Journal of Computational Chemistry**, v. 30, n. 16, p. 2785–2791, 2009.
- MURAMATSU, T. et al. SARS-CoV 3CL protease cleaves its C-terminal autoprocessing site by novel subsite cooperativity. **Proceedings of the National Academy of Sciences**, v. 113, n. 46, p. 12997–13002, 2016.

- O'BOYLE, N. M. et al. Open Babel: An open chemical toolbox. **Journal of cheminformatics**, v. 3, n. 1, p. 1–14, 2011.
- OU-YANG, S. S. et al. **Computational drug discovery. Acta Pharmacologica Sinica**, 27 set. 2012.
- PAINTER, W. P. et al. Human safety, tolerability, and pharmacokinetics of molnupiravir, a novel broad-spectrum oral antiviral agent with activity against SARS-CoV-2. **Antimicrobial Agents and Chemotherapy**, v. 65, n. 5, 1 maio 2021.
- PARRINELLO, M.; RAHMAN, A. Polymorphic transitions in single crystals: A new molecular dynamics method. **Journal of Applied physics**, v. 52, n. 12, p. 7182–7190, 1981.
- PELLEGRINI, C.; STÖHR, J. X-ray free-electron lasers - Principles, properties and applications. **Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment**, v. 500, n. 1–3, p. 33–40, 3 nov. 2003.
- PILLAIYAR, T. et al. **A medicinal chemistry perspective of drug repositioning: Recent advances and challenges in drug discovery. European Journal of Medicinal Chemistry** Elsevier Masson s.r.l., , 1 jun. 2020.
- RAJ, R. Analysis of non-structural proteins, NSPs of SARS-CoV-2 as targets for computational drug designing. **Biochemistry and Biophysics Reports**, v. 25, p. 100847, mar. 2021.
- ROBSON, F. et al. **Coronavirus RNA Proofreading: Molecular Basis and Therapeutic Targeting. Molecular Cell** Cell Press, , 3 set. 2020.
- SCHLICHTING, I.; MIAO, J. **Emerging opportunities in structural biology with X-ray free-electron lasers. Current Opinion in Structural Biology**, out. 2012.
- SHANG, J. et al. Cell entry mechanisms of SARS-CoV-2. **Proceedings of the National Academy of Sciences of the United States of America**, v. 117, n. 21, 2020.
- SILVERMAN, R. B.; HOLLADAY, M. W. *The Organic Chemistry of Drug Design and Drug Action*. Elsevier. **Burlington, MA**, v. 530, 2004.
- SINGH, A. K. et al. **Molnupiravir in COVID-19: A systematic review of literaturef. Diabetes and Metabolic Syndrome: Clinical Research and Reviews** Elsevier Ltd, , 1 nov. 2021.
- SINGHAL, T. **A Review of Coronavirus Disease-2019 (COVID-19). Indian Journal of Pediatrics** Springer, , 1 abr. 2020.

- SOLA, I. et al. **Continuous and Discontinuous RNA Synthesis in Coronaviruses. Annual Review of Virology** Annual Reviews Inc., , 9 nov. 2015.
- SOUSA DA SILVA, A. W.; VRANKEN, W. F. ACPYPE-Antechamber python parser interface. **BMC research notes**, v. 5, p. 1–8, 2012.
- STICHER, Z. M. et al. Analysis of the Potential for N 4-Hydroxycytidine To Inhibit Mitochondrial Replication and Function. 2020.
- SU, H. et al. **Molecular Insights into Small-Molecule Drug Discovery for SARS-CoV-2. Angewandte Chemie - International Edition** John Wiley and Sons Inc, , 26 abr. 2021.
- SYED, Y. Y. Molnupiravir: First Approval. **Drugs**, v. 82, n. 4, p. 455–460, 1 mar. 2022.
- TURNER, P. J. XMGRACE, Version 5.1. 19. **Center for Coastal and Land-Margin Research, Oregon Graduate Institute of Science and Technology, Beaverton, OR**, v. 2, 2005.
- VAN GUNSTEREN, W. F. et al. Biomolecular simulation: the GROMOS96 manual and user guide. **Vdf Hochschulverlag AG an der ETH Zürich, Zürich**, v. 86, p. 1–1044, 1996.
- 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, R. et al. Decoding SARS-CoV-2 Transmission and Evolution and Ramifications for COVID-19 Diagnosis, Vaccine, and Medicine. **Journal of Chemical Information and Modeling**, v. 60, n. 12, p. 5853–5865, 2020.
- WU, C. et al. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. **Science**, v. 3, n. 3, p. 1–8, 2020.
- XU, X. et al. Molecular model of SARS coronavirus polymerase: Implications for biochemical functions and drug design. **Nucleic Acids Research**, v. 31, n. 24, p. 7117–7130, 15 dez. 2003.
- ZHOU, S. et al.  $\beta$ -d-N4-hydroxycytidine Inhibits SARS-CoV-2 through lethal mutagenesis but is also mutagenic to mammalian cells. **Journal of Infectious Diseases**, v. 224, n. 3, p. 415–419, 1 ago. 2021.