Investigation of potential inhibitors of nsp15 for COVID-19 treatment

Project Thesis submitted in partial fulfillment for the requirement for the degree of

Bachelor of Technology

in

Bioinformatics

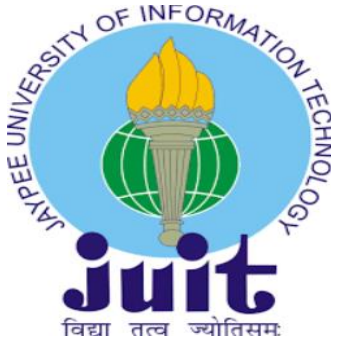
Submitted by –

Arundhati Mehta

181502

Under the guidance of

Dr. Raj Kumar



Department of Bio – Informatics and Bio – Technology

Jaypee University of Information Technology

Waknaghat, Solan, Himachal Pradesh – 173234

Content

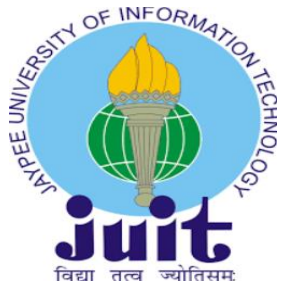
CANDIDATE'S DECLARATION

I hereby declare that the work presented in this project titled "Investigation of potential inhibitors of nsp15 for COVID-19 treatment" in partial fulfillment of the requirements for the award of the degree of Bachelor of Technology in Bioinformatics submitted in the department of Biotechnology and Bioinformatics, Jaypee University and Information Technology, Waknaghat is an authentic record of our work carried from August 2021 to May 2022 under the supervision of Dr. Raj Kumar (Assistant Professor (Grade II)) Department of Biotechnology and Bioinformatics.

Arundhati Mehta
181502

This is to certify that the above statement made by the candidate is true to the best of my knowledge.

Dr. Raj Kumar
Assistant Professor (Grade)
Department of Biotechnology and Bioinformatics



Certificate

This is to certify that the work titled **"Investigation of potential inhibitors of nsp15 for COVID-19 treatment"** submitted by Arundhati Mehta (181502) in partial fulfillment for the award of the degree of B. Tech. in Bioinormatics of Jaypee University of Information Technology, Solan has been carried out under my supervision. This work has not been submitted partially or wholly to any other University or Institute for the award for this or any other degree or diploma.

Dr. Raj Kumar

Assisstant Professor (Grade II)

"Department of Biotechnology and Bioinformatics

Jaypee University of Information Technology (JUIT)

Waknaghat, Solan, India – 173234"

Acknowledgement

I would like to express my sincere gratitude to my supervisor Dr. Raj Kumar for his invaluable guidance, constant supervision and suggestions throughout the course of this project. I would like to thank him for introducing me to the subject of CADD and Structural Biology and motivating me to explore this domain and apply this knowledge to this project.

I thank Dr. Sudhir Kumar, Head, of the Department of Biotechnology and Bioinformatics, for providing me the opportunity and facilities to carry out the project and for guiding and motivating me whenever required.

It is my privilege to thank my parents for their constant motivation and encouragement. They were always there to provide an affectionate shoulder at odd times or when unsatisfactory results led to frustrations. I am and will always be indebted to their support and care.

Arundhati Mehta

181502

List of figures

Figure22. 2D structure of Nirmatrelvir and Erinacine-A

List of Tables

Abstract

"Severe Acute Respiratory Syndrome coronavirus 2 (SARS-CoV-2) has caused a global pandemic." To manage the illness and spread of the virus, the development of a safe and effective antiviral is critical. Non- structural viral proteins are a promising target for therapeutic development. Non-strcutural protein 15 (Nsp15) has proven to be one of the most essential non-structural proteins in the SARS-CoV-2 genome as it plays a vital role in the replication process of the coronavirus. Medicinal fungi have a lot of secondary metabolites, which are an important and diversified chemical space of natural goods. Medicinal fungi has been used to cure human illnesses in traditional medicine from many years. Here, I have selected 10 secondary mushroom metabolites that have been virtually screened with nsp15 macromolecule to find its potential inhibitors. Three compounds were chosen as reference inhibitors for the present study. The binding energy scores of reference inhibitors nirmatrelvir, remdesivir and ritonavir were -8.3 kcal/mol, -8.2 kcal/mol, -7.3 kcal/mol. Further the compound Erinacine-A exhibited lower binding energy score and better molecular interactions then the known reference inhibitors. Therefore, we propose that Erinacine-A is a potential inhibitor of SARS-CoV-2 Nsp15 and may be considered for further pre-clinical studies.

Chapter 1

Introduction

Introduction

“The pandemic caused by SARS-CoV-2 is not over yet, despite all the efforts from the health and scientific community. Vaccination is a pivotal method to scrimmage this virus. However, we still urge the development of the antivirals to reduce the graveness of and the progression of the COVID-19 disease. For this, a deep understanding of the mechanisms involved in the viral replication is necessary. Coronaviruses are important pathogens of animals and human with high zoonotic potential. Fungi have a lot of secondary metabolites, which are an important and diversified chemical space of natural goods. Nsp15 is a uridine specific endoribonuclease which is conserved in the coronavirus family. They have shown common traits of hexameric assembly. Medicinal fungi have been used to cure human illnesses in traditional medicine from many years. Such ten Secondary Mushroom metabolites were selected and a library of was prepared. Only those metabolites with medicinal therapeutic properties were observed were chosen. Two metabolites with antiviral poperties (Ganodermadiol and Lucidadiol), four metabolites with anti-allergic properties (Cyathatriol, Neosarcodonin, Erinacine, Sarcodonin), while others with anti-inflammatory properties were chosen.

Chapter 2

Review of Literature

Introduction

The COVID-19 pandemic is an ongoing global pandemic of coronavirus disease 2019 which is causedd by severe acute respiratory syndrome coronvirus 2 (SARS-CoV-2). It was first reported in Wuhan, China, in December 2019. The World Health Organisation (WHO) declared a Public Health Emergency of Interntional Concern on 30th January 2020 and a pandemic on 11th March 2020.[1]

The Severe acute respiratory sysndrome (SARS) is a viral respiratory disease caused by SARS-associated coronavirus. The SARS-CoV-2 virus infects a wide range of cells and systems of body. It affects the upper respiratory tract – the sinuses, nose, throat and the lower respiratory tract – the windpipe and lungs.[2] The lungs are most affected by virus because it accesses host cells via the receptor for the enzyme angiotensin – converting enzyme 2 (ACE2), which is most abundant on the surface of type 2 alveolar cells of the lungs.

The most common symptoms include mild symptoms include mild symptoms such as dry cough, fatigue and fever. Approximately 15% progress to severe pneumonia and 5% develop acute respiratory distress syndrome, multiple organ failure and even death.

Structure of SARS-CoV-2

Each SARS-CoV-2 virion is 50-200 nanometers in diameter. Like other coronaviruses it has four structural proteins known as S (spike) protein, E (envelope) protein, M (membrane) protein and N (nucleocapsid) protein. The N protein holds the RNA genome and the S, E, M protein create the viral envelope.[3]

The pathogenesis of Covid – 19 is believed to be driven from two forces. The early stage of infection progression, which is driven by fusion, entry, replication of SARS – CoV – 2. This is also known as replication cycle, which is modulated by viral proteins. In the late stage of infection spreading, it is driven by a tremendous inflammatory and immune response to SARS – CoV – 2 which results in high tissue damage.[4] Then, both the proteins of the virus and the host factors are essential for the pathogenesis of Covid – 19.

The therapeutic targets of Covid – 19 are classified into target virus and target host. An analysis of the therapeutic targets is also being done in continuation based on the viral and host factors which occurs on the levels of DNA, RNA and proteins which involves both classic and both novel signalling pathways and that even comprises of the promising epigenetic mechanisms, which would then contribute to SARS – CoV – 2.[5]

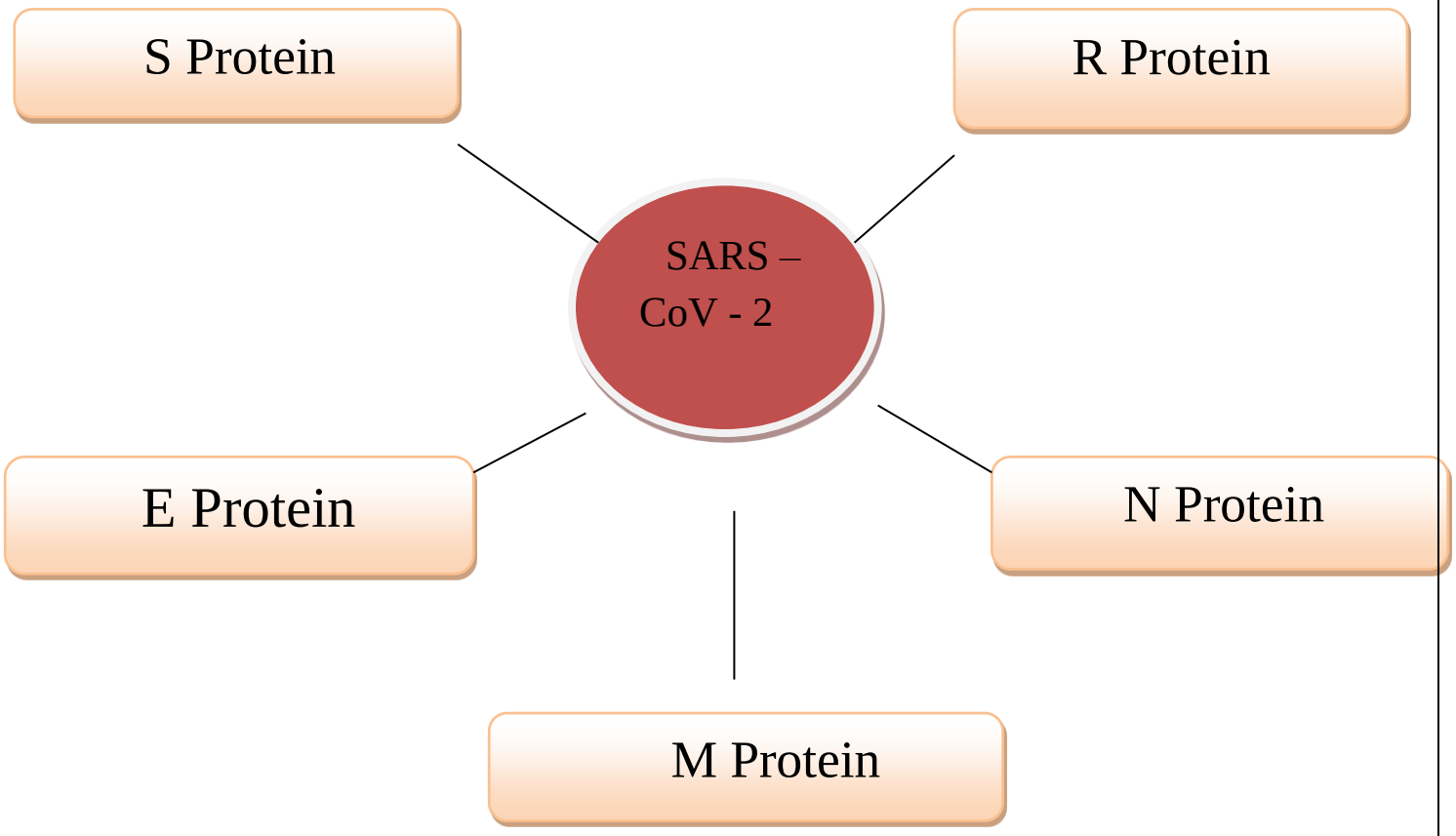


Fig2. Different proteins of SARS-CoV-2

NSP

NSP (Non-Structural Protein) is a protein which is encoded by a virus which is not a part of the viral particle. The non-structural proteins are involved in impeding inborn immunity and also in inducing virus replication. The nsps are important parts of the RTC – replication and transcription complex and immune system evasion. The non-structural proteins help the virus in forming the replication and transcription complex.[6]

Nsp1

Nsp1 is an important virulent factor on the virus-host interaction surface and it has multiple effects on that surface such as inhibiting host mRNA tranlation andd prevailing infffflfammatory cytokkines and chemokines. The nsp1 are mostly conserved in β – CoVs and α – CoVs. After researching and experimenting, it has been reported that nsp1 binds to 40S ribosomal subunit and then inhibits host mRNA translation and then it subdues the translation of capped cellular mRNA through blockade of steps involved in 48S and 80S initiaation complex formation. Other important activities of nsp1 are adjustment in the cell cycle and biomolecular nuclear-cytoplasmic transport. [7]

Number of amino acids – 180

Molecular weight – 19775.31 kDa

Nsp2

Nsp2 is also an important virulent factor and an RNA binding protein that accumulates in viroplasms which actually are cytoplasmic insertions. It is involved in genome replication. Along with the nsp1, nsp2 is involved in suppressing of the gene expression. Through the experiments, it has been stated that the nsp2 is either functionally not essential for viral replication or the proteins retains its function in an uncleaved form. Other important activities of nsp2 are cell cycle progression, apoptosis, mitochondrial biogenesis. [8]

Number of amino acids – 638

Molecular weight - 70.52 kDa

Nsp3

Nsp3 is a papain like proteinase protein. It is the largest protein encoded by the coronaviruses. It is a very important component of the replication and transcription process. Nsp3 has many conserved domains because of the long sequense it possesses such as sRNA binding, papain like protease and transmembrane domains. The inhibition of nsp3 protease activity is an imporatant target for antiviral activity. [9]

Number of amino acids – 1945

Molecular weight – 200 kDa

Nsp4

Nsp4 is a membrane-spanning protein which also functions in anchoring the viral RTC which is viral replication and transcription process to modified endoplasmic reiculum membranes. It interacts with nsp3 and then host proteins which plays an important role in membrane rearrangement in SARS CoV. Just like nsp3, nsp4 also has transmembrane domains.[10]

Number of amino acids – 500

Molecular weight – 56.2 kDa

Nsp5

Nsp5 is the main protease which plays an important role in the formation of a polyprotein cluster which then translates viral RNA and also in the post-translational modification of viral proteins. This formation usually takes place with the help of ADP ribose phosphatase activity. It is a 3C like proteinase. Nsp5 is involved in lung inflammation process through its connections with the Histone Deacetylase2.[11]

Number of amino acids – 306

Molecular weight – 30-34 kDa

Nsp6

Nsp6 is putative transmembrane domain protein. It generates autophagosomes from the ER – Endoplasmic Reticulum and it inhibits the lysosome – autophagy system. Autophagosomes helps in assembly of replicase proteins. Nsp6 prevents expansion of the autophagosome and the lysosomes which, in turn, prevents autophagosomes from delivering viral components for degradation in the lysosomes.[11]

Number of amino acids – 290

Molecular weight – 33 kDa

Nsp7andNsp8

Nsp8 is a cofactor of a peptide which makes a heterodimer with the Nsp7. This heterodimer of Nsp 7 and 8 makes a complex with the Nsp12. This complex of Nsp7, Nsp8, Nsp12 helps in formation of RNA polymerase complex if a monomer of Nsp8 also joins in the complex. Nsp8 is encoded on the open reading frame ORF1a. It has two sub domains. The first domain has an α-turn-α motif. The second domain has four antiparallel β-strands which consists of C-terminal subdomain which has also an α3 inserted in it.[12]

Number of amino acids in Nsp7 – 83

Molecular weight of Nsp7 – 12.3 kDa

Number of amino acids in Nsp8 – 198

Molecular weight of Nsp8 – 21.89 kDa

Nsp9

Nsp9 is a single stranded RNA binding subunit. It has a diverse form of dimerization that promotes the biological function. It forms a dimer from a conserved region which is an α-helical motif, where the interruption of key residues within this region reduces RNA bindng and SARS-CoV poliferation. The arrangment of monomers with the dimers of Nsp9 is very well conserved at

different CoVs. It has also been stated that Nsp9 has been found interacting with the DEAD-box RNA helicase 5 which is a cellular protein.[13]

Number of amino acids – 113

Molecular weight – 15.5 kDa

Nsp10

Nsp10 is a subunit which consists of two subdomains – one is a helical α-subdomain composed of two anti-parallel β sheets, second is a short helix α5 and several coiled coil regions. Nsp10 has two zinc binding sites. One site is formed by three cysteine residues (Cys74, Cys77, Cys90). Another site is formed by four cysteine residues (Cys117, Cys120, Cys128, Cys130). On the former site a histidine residue is located α2 and α3 and this stabilizes them.[14][15]

Number of amino acids –139

Molecular weight – 16 kDa

Nsp11

Nsp11 is made up of 13 amino acids out of which first nine amino acids are identical to the first nine amino acids of Nsp12. According to the researches and experimental data, it is stated that in the presence of membrane mimetic environment and negatively charged and neutral liposomes, nsp11 remains distorted. [16][17]

Number of amino acids – 13

Molecular weight – 4.8 kDa

Nsp12

Nsp12 is an RNA dependent RNA polymerase. It copies viral RNA. As the RNA dependent RNA polymerase, Nsp12 helps in synthesizing viral DNA. The structure of Nsp12 consists of an RdRp domain and a nidovirus RdRp associated NiRAN domain which is nucleotidyltransferase domain. Through experimental data it has been revealed that Nsp12 has also shown the ability to suppress interferon expression. It has also been found that it could also suppress a receptor kinase that

triggers the retinoic acid-inducible gene and melanoma differentiation-associated protein signalling pathways and that mobilizes interferon beta production.[18]

Number of amino acids – 932

Molecular weight – 107.9 kDa

Nsp13

Nsp13 is a good target for anti-virals as it has high sequence conservation and has an essential role in viral replication. Nsp13 belongs to the helicase superfamily. According to the experimental data, Nsp13 has shown interactions with the viral RNA dependent RNA polymerase i.e Nsp12. Nsp13 possesses RNA5' triphosphatase activity which plays an important role in the formation of viral 5' mRNA Nsp13 consists of five domains – a N-terminal ZBD i.e a Zinc binding domain, a helical stalk domain, a beta barrel domain, two RecA like helicase subdomains which also contains the residues which are responsible for nucleotide binding and hydrolysis.[18]

Number of amino acids – 601

Molecular weight – 69.15 kDa

Nsp14

Nsp14 exon is known to have 3'-5' exoribonuclease activity as well as N7-methyltransferase activity. The exoribonuclease activity is important for the maintenance of the large RNA genome and the methyltransferase activity is important for the stability of the viral RNA. The exon domain consists of a central, twisted β sheet with the exception of β3 and are strenthened by α-helices on each side.[19]

Number of amino acids – 527

Molecular weight – 59.8 kDa

Nsp16

Nsp16 is a 2'-0-methyltransferase (2'-0-Mtase) which forms a part of the replication – transcription complex. It plays a very important role in immune evasion by mimicking human. The

nsp16 consists of twelve β-strands, seven α-helices and five 3_{10} helices. The core of the structure is actually made up of strands β1-β7 and helices αa, αz, αd and αe. [17]

Number of amino acids – 298

Molecular weight – 16 kDa

Nsp15

Nsp15 is a uridine specific endoribonuclease. It is a conserved protein in the family of coronaviruses i.e., Coronaviridae. Coronavirus is a member of the Nidovirus family. The Nidovirus encloses RNA virus which is a positive – strand virus and is very large. The size of this virus ranges from 12 to 41 kb. Nsp15 is mainly, on the whole, conserved across the Coronaviridae family. [20] Many crystal structures of Nsp15 as identified and analysed via different procedures have shown a common hexameric assembly. In many experimental data it was stated that Nsp15 concessionary cleaves RNA substrates 3' of uridines. This contributes to a reason why Nsp15 is also known as endoU which insinuates its cleavage specificity. The endonuclease function of Nsp15 is very important for its viral life cycle. Another important role that Nsp15 plays is that its nuclease activity is very crucial for evading the activation of hosts immune responses. [21]

Structure of Nsp15

"The Nsp15 trimers make a common hexameric assembly of a dimer. A protomer of Nsp15 is made up of three domains –

- an N terminal domain which helps in oligomerization,
- a middle domain,
- a C – terminal endonuclease (endoU) domain which shares homology with other endoU enzymes."

The active site of Nsp15 contains two histidine residues that conserved inside that active site. These residues are important for the catalysis of the active site of Rnase A. Rnase A catalyzes a reaction that includes two steps – initially, it generates a 2'3' – cyclic phosphate and then it is

hydrolysed which helps in formation of 3'-phosphate. [20] The hexameric assembly of Nsp15 contains six protomers which has a D3 symmetry. In Nsp15, the active sites lie in protomers individually which is near the interface with the endoU domain. The allosteric communication between the protomers of Nsp15 is a result of the positions of the active sites with the neighbouring protomer. [21]

For the nuclease activity in the active sites, the active sites contain three residues which also includes a lysine residue and a histine residue and corresponds to the active sites of Rnase A.

Two functions are performed within the Rnase A active site –

- one is functioning of histidine as a base for the activation of 2'OH and
- other is functioning of histidine as an acid to donate a proton to 5'OH.[22]

FDA approved drugs against SARS-CoV-2

1. Nirmatrelvir [23]
   PubChem CID – 155903259

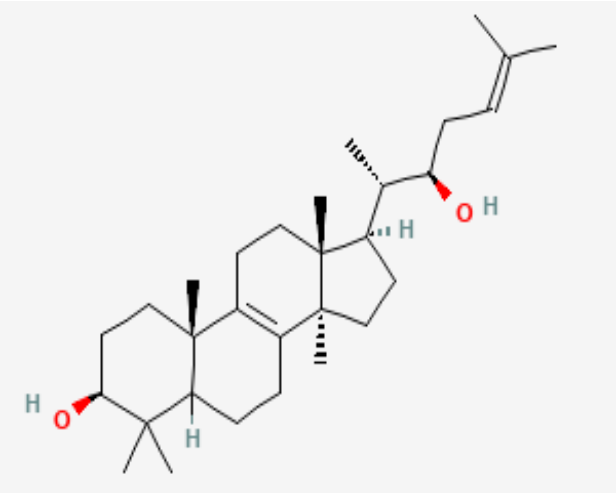


Fig3. 2D structure of Nirmtrelvir

Molecular Formula - $C_{23}H_{32}F_3N_5O_4$
Molecular Weight - 499.0
"IUPAC -
(1R,2S,5S)-N-[(1S)-1-cyano-2-[(3S)-2-oxopyrrolidin-3-yl]ethyl]-3-[(2S)-3,3-dimethyl-2-[(2,2,2-trifluoroacetyl)amino]butanoyl]-6,6-dimethyl-3-azabicyclo[3.1.0]hexane-2-carboxamide"

Nirmatrelvir is an orally bioavailable inhibitor of the severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) 3CL protease (3CLpro), which performs antiviral activity against SARS-CoV-2 and other coronaviruses. Nirmatrelvir targets the activity of SARS-CoV-2 3CLpro. This inhibits the proteolytic cleavage of viral polyproteins, which further inhibits the formation of viral proteins including helicase, single-stranded-RNA-binding protein, RNA-dependent RNA polymerase, 20-O-ribose methyltransferase, endoribonuclease and exoribonuclease. This prevents viral transcription and replication.

2. Remdesivir [24]

   PubChem CID – 121304016

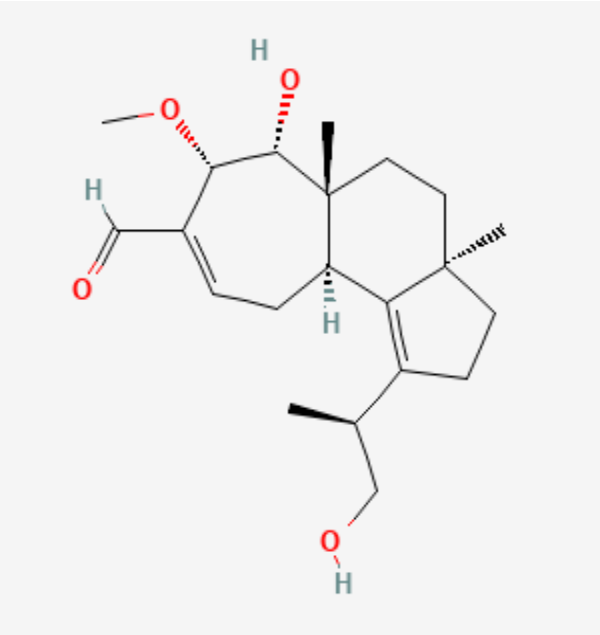


Fig 4. 2D Structure of Remdesivir

Molecular Formula - $C_{27}H_{35}N_6O_8P$

Molecular Weight - 602.6

"IUPAC -

2-ethylbutyl(2S)-2-[[[(2R,3S,4R,5R)-5-(4-aminopyrrolo[2,1-f][1,2,4]triazin-7-yl)-5-cyano-3,4-dihydroxyoxolan-2-yl]methoxy-phenoxyphosphoryl]amino]propanoate"

Remdesivir is an antiviral nucleotide used for therapy of severe novel coronavirus disease 2019 (COVID-19) caused by SARS CoV-2 infection. Remdesivir therapy is for 3 to 10 days and is accompanied by reversible mild-to-moderate elevations in serum aminotransferase levels.

3. **Ritonavir** [25]

   PubChem CID – 392622

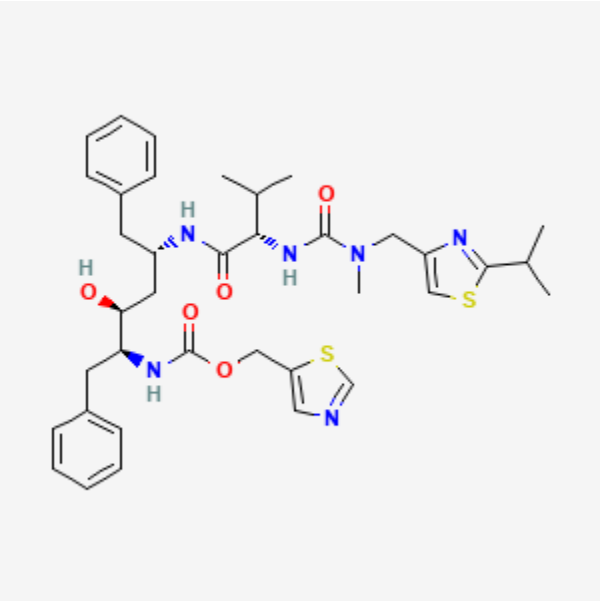


Fig5. 2D Structure of Ritonavir

Molecular Formula - $C_{37}H_{48}N_6O_5S_2$

Molecular Weight - 720.9

"**IUPAC Name** -

1,3-thiazol-5-ylmethylN-[(2S,3S,5S)-3-hydroxy-5-[[(2S)-3-methyl-2-[[methyl-[(2-propan-2-yl-1,3-thiazol-4-yl)methyl]carbamoyl]amino]butanoyl]amino]-1,6-diphenylhexan-2-yl]carbamate"

Ritonavir is an antiretroviral protease inhibitor that is widely used in combination with other protease inhibitors in the therapy and prevention of human immunodeficiency virus (HIV) infection and the acquired immunodeficiency syndrome (AIDS). Ritonavir can cause asymptomatic elevations in serum aminotransferase levels and sometimes can lead to clinically apparent acute liver injury. In HBV or HCV coinfected patients, highly active antiretroviral therapy with ritonavir may result of an exacerbation of the underlying chronic hepatitis B or C.

Secondary Structure of Mushroom Metabolites

1. **Inotodiol**

PubChem CID - 182264

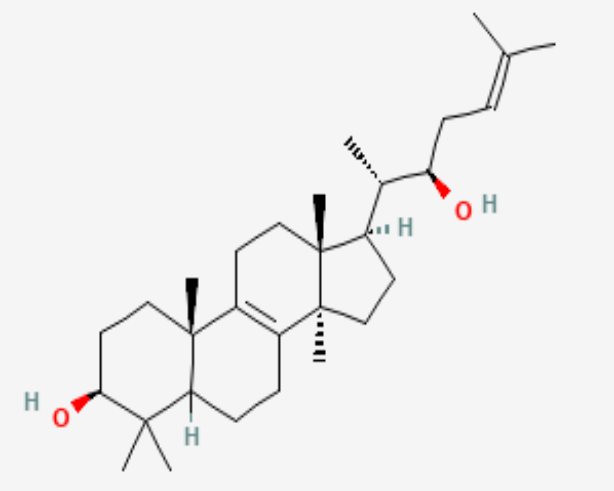


Fig6. 2D Structure of Inotodiol

Inotodiol is an anti-inflammatory sterol isolated from Inotonus obliquus.

Molecular Formula - $C_{30}H_{50}O_2$

Molecular weight – 442.7

"IUPAC Name – (3*S*,5,10,13,14,17)-17-[(2S,3)-3-hydroxy-6-methyhept-5-en-2-yl]-4,4,10,31,14-pentmethy-2,3,6,7,11,12,15,16,17-decahydro-1*H*-cyclolpents[a]phenanthren-3-ol"

Hydrogen bond donor and acceptor count – 2 and 2

Exact Mass – 442.381080833

InChl –

InChl = 1S/C30H50O2/c1-19(2)9-11-24(31)20(3)21-13-17-30(8)23-10-12-25-27(4,5)26(32)15-16-28(25,6)22(23)14-18-29(21,30)7/h9,20-21,24-26,31-32H,10-18H2,1-8H3/t20-,21+,24+,25,26-,28+,29+,30-/m0/s1

InChl Key - KKWJCGCIAHLFNE-UJHWODAZSA-N

"Canonical SMILES -

CC(C1CCC2(C1(CCC3=C2CCC4C3(CCC(C4(C)C)O)C)C)C)C(CC=C(C)C)O"

"Isomeric SMILES –

C[C@@H]([C@H]1CC[C@@]2([C@@]1(CCC3=C2CCC4[C@@]3(CC[C@@H](C4(C)C)O)

C)C)C)[C@@H](CC=C(C)C)O"

2. **<u>Neosarcodonin A</u>**

PubChem CID - 101153516

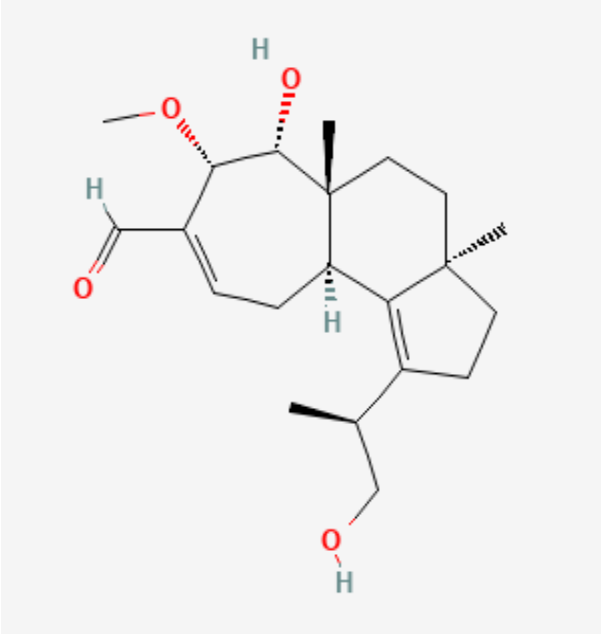


Fig7. 2D Structure of Neosarcodonin A

Neosarcodonin A is an anti-inflammatory cyathanediterpenoids from *Sarcodonscabrosus*.

Molecular Formula - $C_{21}H_{32}O_4$

“IUPAC name -

(3aR,5aR,6R,7S,10aR)-6-hydroxy-1-[(2S)-1-hydroxypropan-2-yl]-7-methoxy-3a,5a-dimethyl-2,3,4,5,6,7,10,10a-octahydrocyclohepta[e]indene-8-carbaldehyde”

Molecular Weight - 348.5

Hydrogen Bond Donor and Acceptor Count - 2 and 4

Exact Mass - 348.23005950

InChl -

InChI=1S/C21H32O4/c1-13(11-22)15-7-8-20(2)9-10-21(3)16(17(15)20)6-5-14(12-23)18(25-4)19(21)24/h5,12-13,16,18-19,22,24H,6-11H2,1-4H3/t13-,16-,18+,19+,20-,21-/m1/s1

InChl Key - KIJQNYNQHIZOJO-MEJVKNBMSA-N

Canonical SMILES - CC(CO)C1=C2C3CC=C(C(C(C3(CCC2(CC1)C)C)O)OC)C=O

“Isomeric SMILES -

C[C@H](CO)C1=C2[C@H]3CC=C([C@@H]([C@@H]([C@@]3(CC[C@]2(CC1)C)C)O)OC)C=O”

3. Cyathatriol

PubChem CID - 101316898

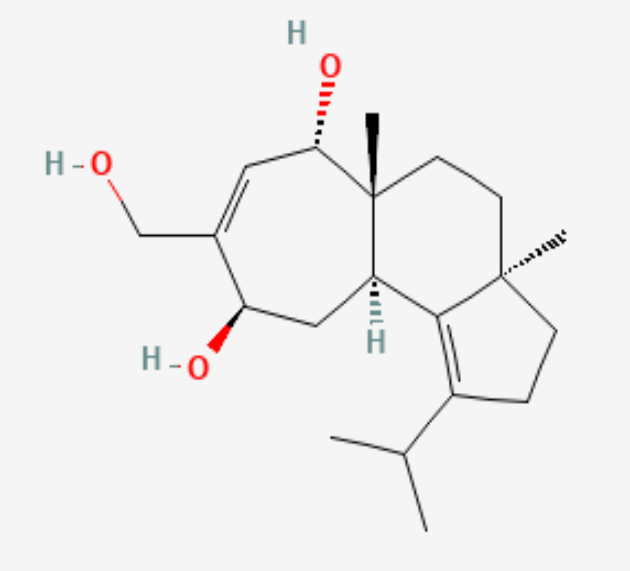


Fig8. 2D Structure of Cyathatriol

Cyathatriol is an anti-inflammatory and cytotoxic cyathanediterpenoid from Cyathusafricanus

Molecular Formula - $C_{20}H_{32}O_3$

"IUPAC name -

(3aR,5aR,6S,9R,10aR)-8-(hydroxymethyl)-3a,5a-dimethyl-1-propan-2-yl-2,3,4,5,6,9,10,10a-octahydrocyclohepta[e]indene-6,9-diol"

Molecular Weight - 320.5

Hydrogen Bond Donor and Acceptor Count - 3 and 3

Exact Mass - 320.23514488

InChl -

InChI=1S/C20H32O3/c1-12(2)14-5-6-19(3)7-8-20(4)15(18(14)19)10-16(22)13(11-21)9-17(20)23/h9,12,15-17,21-23H,5-8,10-11H2,1-4H3/t15-,16-,17+,19-,20-/m1/s1

InChl Key - YQGDZWWLYAMTAU-HPUSYDDDSA-N

Canonical SMILES - CC(C)C1=C2C3CC(C(=CC(C3(CCC2(CC1)C)C)O)CO)O

"Isomeric SMILES –

CC(C)C1=C2[C@H]3C[C@H](C(=C[C@@H]([C@@]3(CC[C@]2(CC1)C)C)O)CO)O"

4. **Cyathin-B3**

PubChem CID - 102117112

.

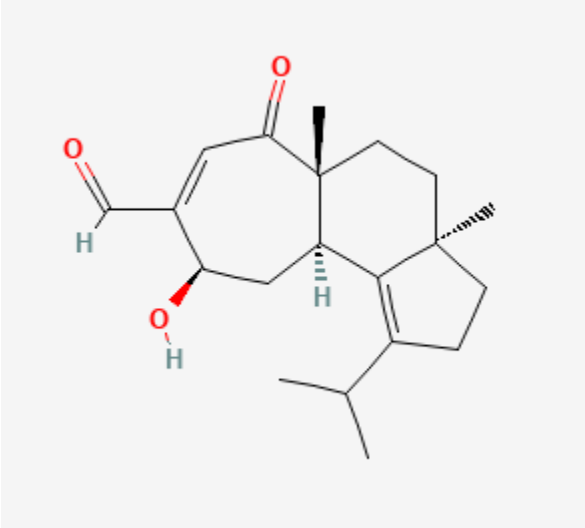


Fig9. 2D Structure of Cyathin-B3

Cyathin-B3 is an anti-inflammatory diterpenoid from *CythusHelenae*

Molecular Formula – $C_{20}H_{28}O_3$

IUPAC name –

(3aR,5aR,9R,10aR)-9-hydroxy-3a,5a-dimethyl-6-oxo-1-propan-2-yl-3,4,5,9,10,10a-hexahydro-2H-cyclohepta[e]indene-8-carbaldehyde

Molecular Weight – 316.4

Hydrogen Bond Donor and Acceptor Count – 1 and 3

Exact Mass – 316.20384475

InChl –

InChI=1S/C20H28O3/c1-12(2)14-5-6-19(3)7-8-20(4)15(18(14)19)10-16(22)13(11-21)9-17(20)23/h9,11-12,15-16,22H,5-8,10H2,1-4H3/t15-,16-,19-,20-/m1/s1

InChl Key – HTEKHSBJKOVLAK-XNFNUYLZSA-N

"Canonical SMILES – CC©C1=C2C3CC(C(=CC(=O)C3(CCC2(CC1)C)C)C=O)O"

"Isomeric SMILES -
CC©C1=C2[C@H]3C[C@H](C(=CC(=O)[C@@]3(CC[C@]2(CC1)C)C)C=O)O"

5. **Erinacine A**

PubChem CID – 10410568

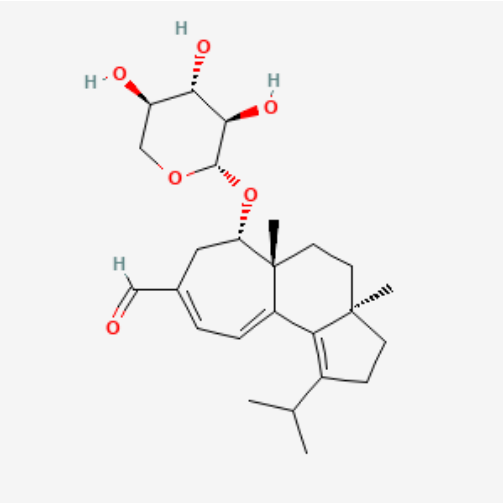


Fig10. 2D Structure of Erinacine

Erinancine A is a therapeutic diterpenoid isolated from *Hericiumerinaceus*.

Molecular Formula - $C_{25}H_{36}O_6$

"IUPAC names-(3aR,5aR,6S)-3a,5a-dimethyl-1-propan-2-yl-6-[(2S,3R,4S,5R)-3,4,5-trihydroxyoxan-2-yl]oxy-2,3,4,5,6,7-hexahydrocyclohepta[e]indene-8-carbaldehyde"

Molecular Weight - 432.5

Hydrogen Bond Donor and Acceptor Count - 3 and 6

Exact Mass - 432.25118886

InChl -

InChI=1S/C25H36O6/c1-14(2)16-7-8-24(3)9-10-25(4)17(20(16)24)6-5-15(12-26)11-19(25)31-23-22(29)21(28)18(27)13-30-23/h5-6,12,14,18-19,21-23,27-29H,7-11,13H2,1-4H3/t18-,19+,21+,22-,23+,24-,25-/m1/s1

InChl Key - LPPCHLAEVDUIIW-NLLUTMDRSA-N

Canonical SMILES -
CC(C)C1=C2C3=CC=C(CC(C3(CCC2(CC1)C)C)OC4C(C(C(CO4)O)O)O)C=O

Isomeric SMILES -
CC(C)C1=C2C3=CC=C(C[C@@H]([C@@]3(CC[C@]2(CC1)C)C)O[C@H]4[C@@H]([C@H]([C@@H](CO4)O)O)O)C=O

6. Lucidadiol

PubChem CID: 10789991

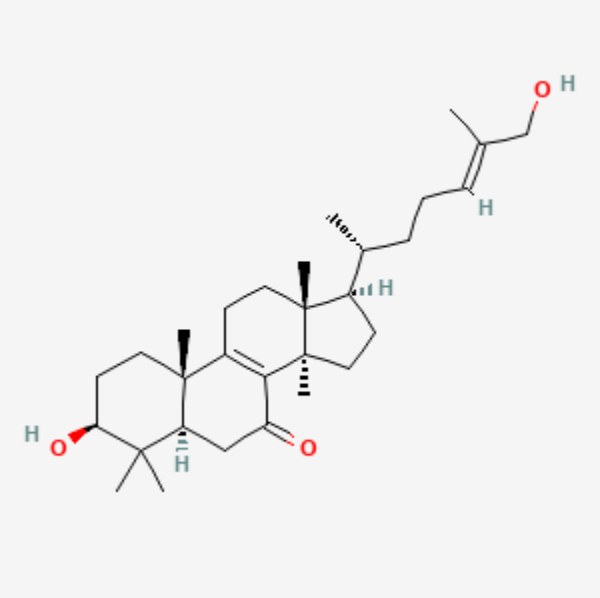


Fig11. 2D Structure of Lucidadiol

Lucidadiol is a natural product found in Ganoderma pfeifferi.

Molecular Formula - $C_{30}H_{48}O_3$

"IUPAC names – (3*S*,5*R*,10*S*,13*R*,14*R*,17*R*)-3-hydroxy-17-[(*E*,2*R*)-7-hydroxy-6-methylhept-5-en-2-yl]-4,4,10,13,14-pentamethyl-1,2,3,5,6,11,12,15,16,17-decahydrocyclopenta[a]phenanthren-7-one"

Molecular Weight – 456.7

Hydrogen Bond Donor and Acceptor Count - 2 and 3

Exact Mass – 456.36034359

InChl –

InChl = 1S/C30H48O3/c1-19(18-31)9-8-10-20(2)21-11-16-30(7)26-22(12-15-29(21,30)6)28(5)14-13-25(33)27(3,4)24(28)17-23(26)32/h9,20-21,24-25,31,33H,8,10-18H2,1-7H3/b19-9+/t20-,21,-24,+25,+28,-29,-30+/m1/s1

InChl Key – AZPOACUDFJKUHJ-GPEQXWBKSA-N

Canonical SMILES –
CC(CCC=C(C)CO)C1CCC2(C1(CCC3=C2C(=O)CC4C3(CCC(C4(C)C)O)C)C)C

Isomeric SMILES – C[C@H0_CC/C=C(\C)/CO)

7. **Enokipodin D**

PubChem CID – 10901419

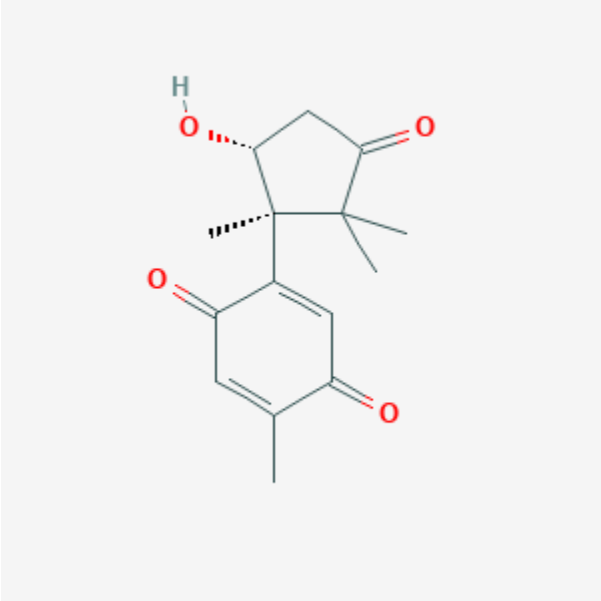


Fig12. 2D Structure of Enokipodin D

Enokipodin D is a natural product found in Flammulina velutipes.

Molecular Formula - $C_{15}H_{18}O_4$

"IUPAC names – 2-[(1S,5R)-5-hydroxy-1,2,2-trimethyl-3-oxocyclopentyl]-5-methylcyclohexa-2,5-diene-1,4-dione"

Molecular Weight – 262.30

Hydrogen Bond Donor and Acceptor Count – 1 and 4

Exact Mass – 262.12050905

"InChl –

InChI=1S/C15H18O4/c1-8-5-11(17)9(6-10(8)16)15(4)13(19)7-12(18)14(15,2)3/h5-6,13,19H,7H2,1-4H3/t13-,15+/m1/s1"

InChl Key – DBTMIHPJDPGOCQ-HIFRSBDPSA-N

Canonical SMILES – CC1=CC(=O)C(=CC1=O)C2(C(CC(=O)C2©C)O)C

Isomeric SMILES – CC1=CC(=O)C(=CC1=O)[C@]2([C@@H](CC(=O)C2©C)O)C

8. **<u>Ganodermadiol</u>**

PubChem CID – 139586903

Molecular Formula – $C_{30}H_{48}O_2$

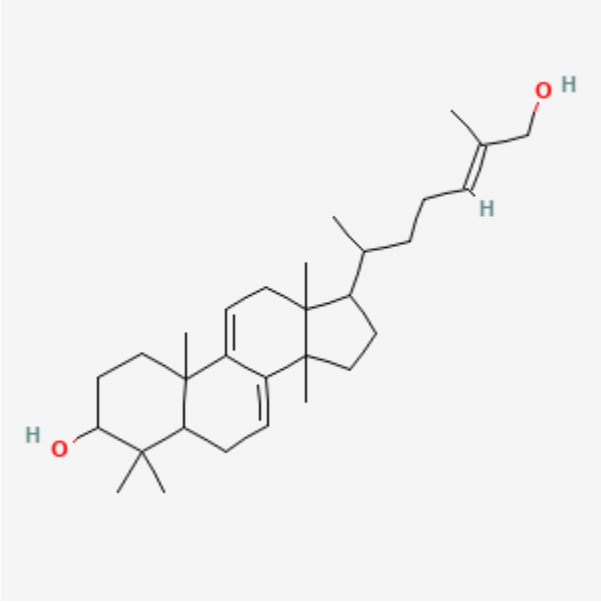


Fig13. 2D Structure of Ganodermadiol

Ganodermadiol is a sterol which was isolated from Ganoderma

IUPAC name - 17-[(*E*)-7-hydroxy-6-methylhept-5-en-2-yl]-4,4,10,13,14-pentamethyl-2,3,5,6,12,15,16,17-octahydro-1*H*-cyclopenta[a]phenanthren-3-ol

Molecular Weight – 440.7

Hydrogen Bond Donor and Acceptor Count – 2 and 2

Exact Mass – 440.365430770

"InChl –

InChI=1S/C30H48O2/c1-20(19-31)9-8-10-21(2)22-13-17-30(7)24-11-12-25-27(3,4)26(32)15-16-28(25,5)23(24)14-18-29(22,30)6/h9,11,14,21-22,25-26,31-32H,8,10,12-13,15-19H2,1-7H3/b20-9+"

InChl Key - AOXXVRDKZLRGTJ-AWQFTUOYSA-N

Canonical SMILES -
CC(CCC=C(C)CO)C1CCC2(C1(CC=C3C2=CCC4C3(CCC(C4(C)C)O)C)C)C

Isomeric SMILES -
CC(CC/C=C(\C)/CO)C1CCC2(C1(CC=C3C2=CCC4C3(CCC(C4(C)C)O)C)C)C

9. Sarcodonin A

PubChem CID: 17747381

Molecular Formula - $C_{20}H_{28}O_3$

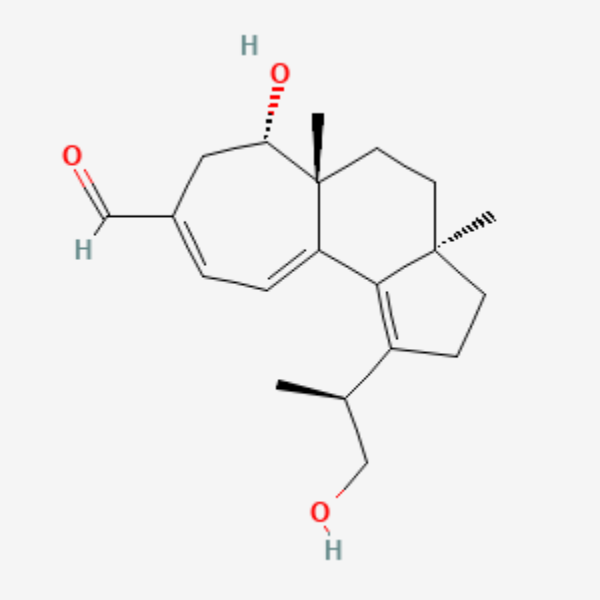


Fig14. 2D Structure of Sarcodonin A

Sarcodonin is a natural product which is found in Sarcodon leucopus.

IUPAC name - (3*aR*,5*aR*,6*S*)-6-hydroxy-1-[(2*S*)-1-hydroxypropan-2-yl]-3*a*,5*a*-dimethyl-2,3,4,5,6,7-hexahydrocyclohepta[e]indene-8-carbaldehyde

Molecular Weight – 316.4

Hydrogen Bond Donor and Acceptor Count – 2 and 3

Exact Mass – 316.20384475

InChl –

InChI=1S/C20H28O3/c1-13(11-21)15-6-7-19(2)8-9-20(3)16(18(15)19)5-4-14(12-22)10-17(20)23/h4-5,12-13,17,21,23H,6-11H2,1-3H3/t13-,17+,19-,20-/m1/s1

InChl Key - HFBBAANNESGPQZ-ISJOWMGUSA-N

Canonical SMILES - CC(CO)C1=C2C3=CC=C(CC(C3(CCC2(CC1)C)C)O)C=O

Isomeric SMILES -
C[C@H](CO)C1=C2C3=CC=C(C[C@@H]([C@@]3(CC[C@]2(CC1)C)C)O)C=O

10. Coprinol

PubChem CID - 42608175

Molecular Formula - $C_{15}H_{22}O_2$

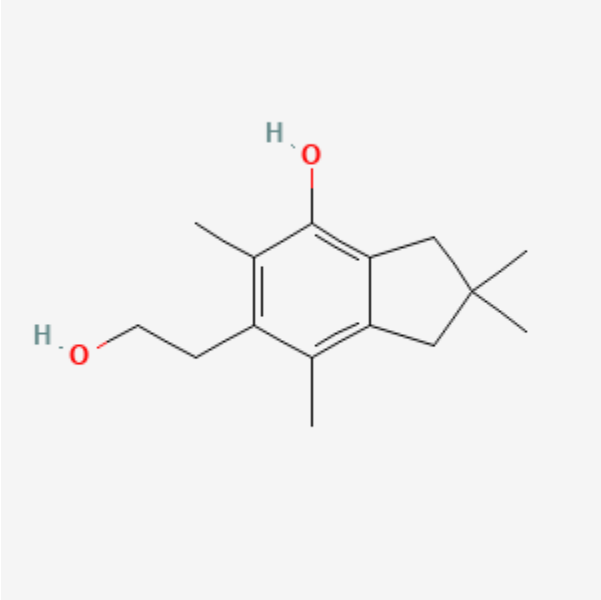


Fig 15. 2D Structure of Coprinol

Coprinol is a natural product found in Coprinus and Coprinopsis cinerea.

IUPAC names - 6-(2-hydroxyethyl)-2,2,5,7-tetramethyl-1,3-dihydroinden-4-ol

Molecular Weight – 234.33

Hydrogen Bond Donor and Acceptor Count – 2 and 2

Exact Mass – 234.161979940

“InChl –

InChI=1S/C15H22O2/c1-9-11(5-6-16)10(2)14(17)13-8-15(3,4)7-12(9)13/h16-17H,5-8H2,1-4H3”

InChl Key - GCMUHPCLXBXQDH-UHFFFAOYSA-N

Canonical SMILES - CC1=C2CC(CC2=C(C(=C1CCO)C)O)(C)C

Chapter 3

Materials and Methodology

Predocking

Protein Preparation

1. **Preparation of the target protein for docking using Discovery Studio -**
   The crytsal structure of Nsp15 endoribonuclease with PDB ID 6vww was downloaded from the RCSB protein data bank. The protein was then prepared for pre – docking and structure-based calculations by converting it into macromolecule. By converting the protein into a macromolecule, residues which were not required for further analysis were highlighted and removed and ligand was obtained.

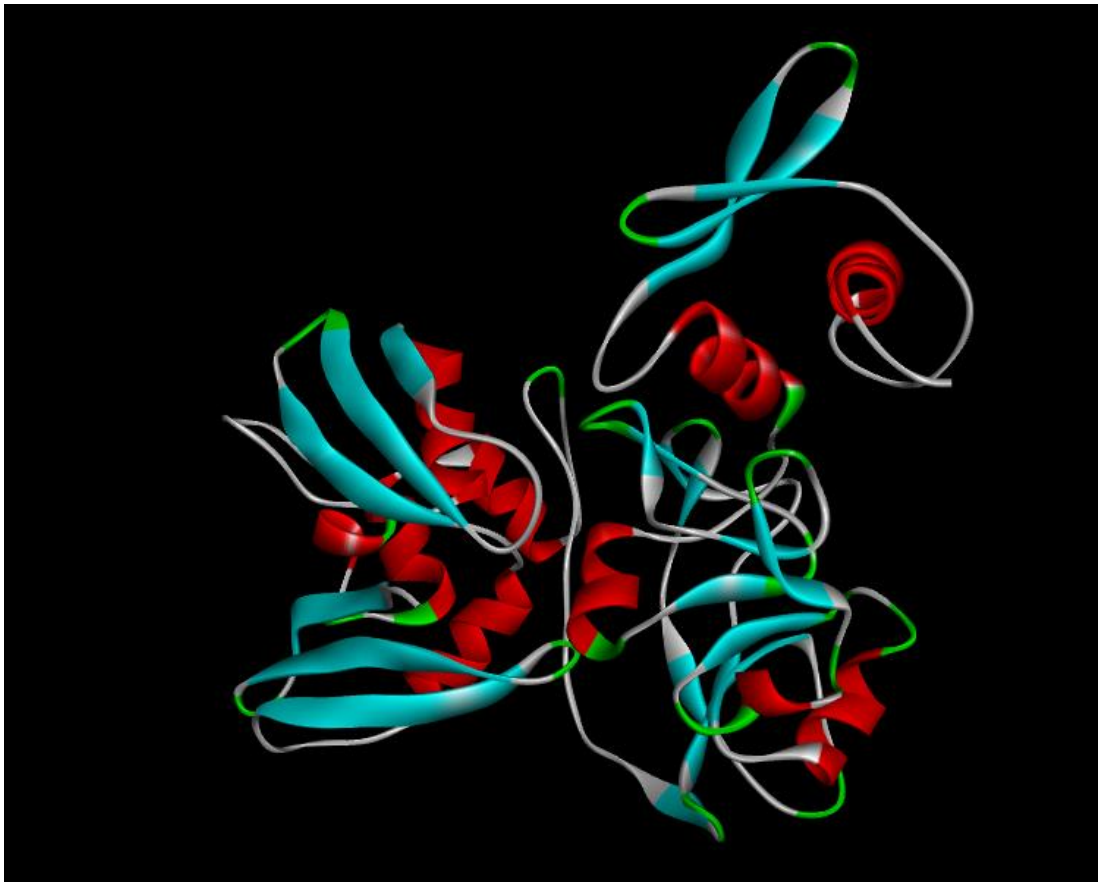


Fig16. Protein preparation for pre-docking

2. **Preparation of the target protein for docking using Chimera** – The obtained ligand was further prepared by removing residues and was minimized using Chimera software which

helped in addition and minimization of missing loops, removal of crystal waters and protonation of the ligand through its algorithms.

3. **Preparation using PyRx –**

   The molecule which was minimized and prepared using Chimera was then loaded on PyRx. The molecule was then minimized and converted into a PDBQT ligand. Through Open Babel tool in PyRx, three drug molecules (ritonavir, remdesivir, nirmatrelvir) were loaded. The drug molecules were also minimized and converted and saved into PDBQT ligands.

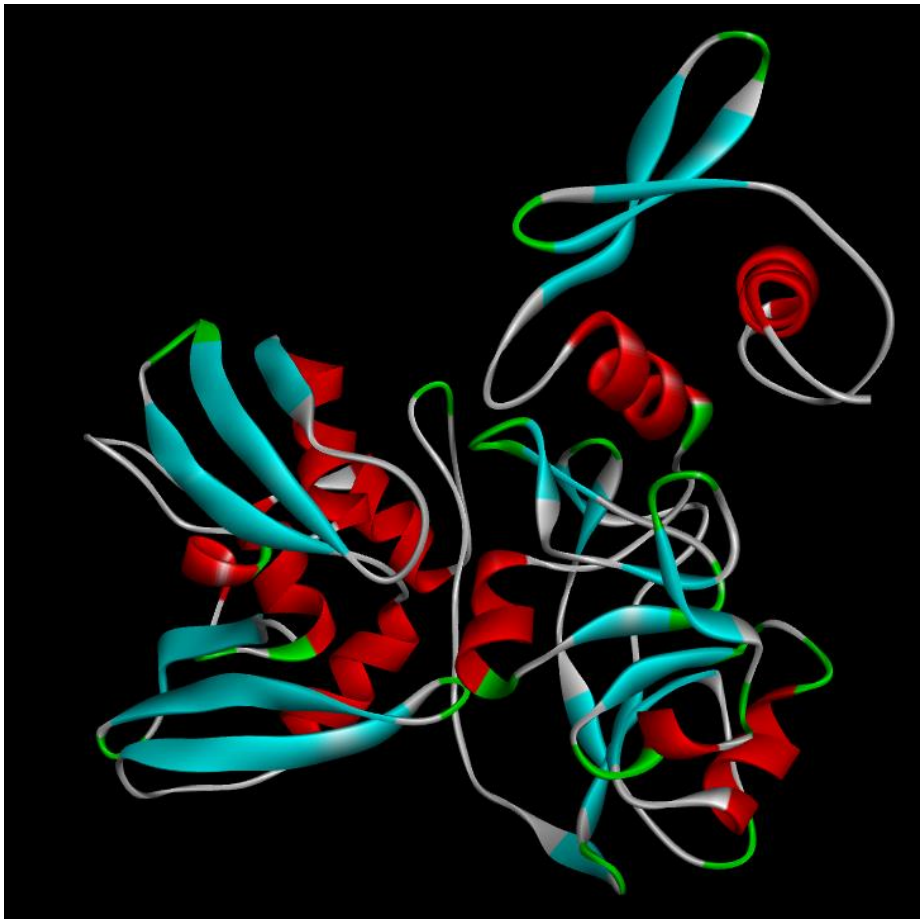


Figure17. Minimized molecule of 6vww protein

4. Using autodock vina wizard, ligand and the three drug molecules were loaded for grid preparation. Exhaustiveness for the docking was set at 8. Coordinates of the grid were set as 25 * 26.86 * 29.61.

5. After the docking of the three drug ligands and the macromolecule experimental and docking pose were calculated by superimposing cocrystals of the endoribonuclease of PDB ID 6vww with drug ligands.

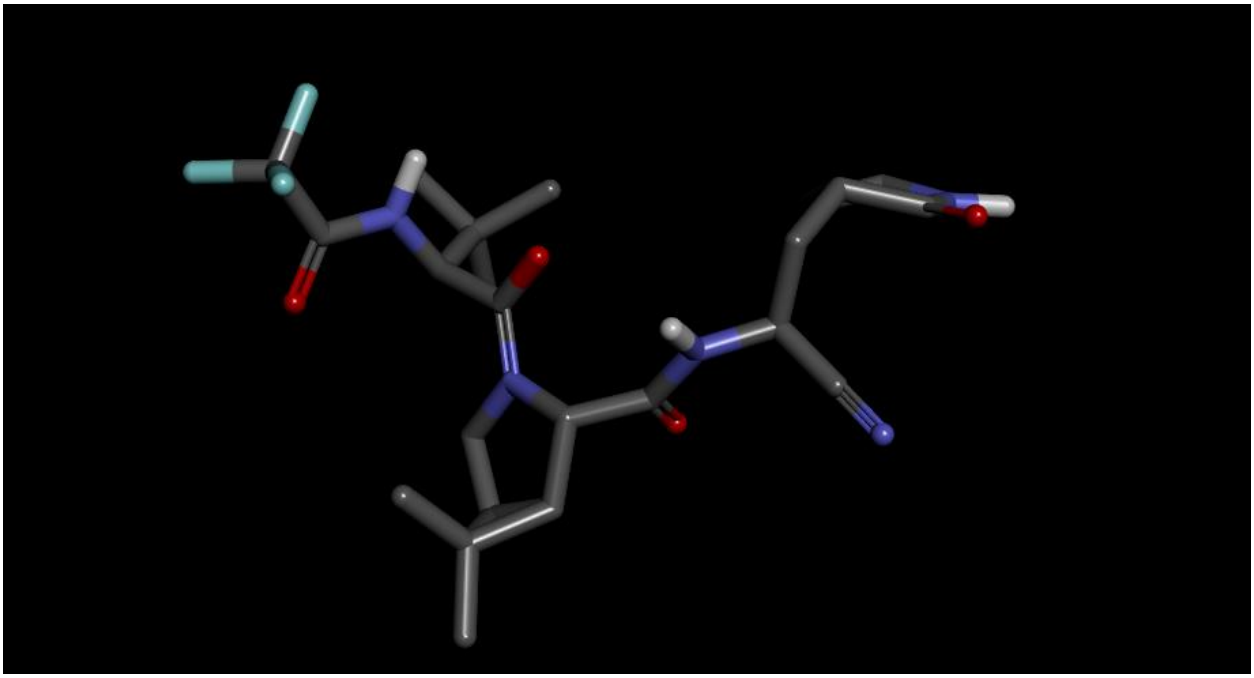


Fig18. Superimposing cocrystals

Secondary mushroom metabolites library

Ten Mushroom metabolites were downloaded from PubChem 3D structures of these secondary metaboloids were saved in SDF and PDBQT file format for further docking procedure. A library of the metabolites was prepared. Only those metabolites were chosen in which medicinal therapeutic properties were seen. Two metabolites with antiviral poperties (Ganodermadiol and Lucidadiol), four metabolites with anti-allergic properties (Cyathatriol, Neosarcodonin, Erinacine, Sarcodonin) while other four metabolites with anti-inflammatory properties were chosen.

Protein preparation and Grid Preparation for Mushroom Metabolites

1. **Preparation of target protein for docking using discovery studio –**

   The crystal structure of Nsp15 endoribonuclease with PDB ID 6vww was downloaded from the RCSB protein data bank. The protein was then prepared for docking by converting it into macromolecule. By converting that protein into macromolecule, residues which were not required for any further analysis were highlighted and removed and ligand was obtained.

2. **Preparation of target protein for docking using Chimera** –

   The obtained ligand was further prepared by removing residues and was minimized using Chimera software, which again helped in addition and minimization of missing loops, removal of crystal waters and protonation of ligands through its algorithms.

3. **Preparation using PyRx –**

   The molecule which was prepared and minimized using Chimera was then loaded in PyRx. The molecule was then converted into a PDBQT ligand. Through Open Babel tool in PyRx, ten mushroom terpenoids (Inotodiol, Erinacine, Lucidadiol, Enokipodin, Sarcodonin, Coprinol, Neosarcodonin, Cyathatriol, Cyathin, Ganodermadiol) were loaded. The terpenoids were also converted and then saved as PDBQT ligands.

4. Using autodock vina wizard, the ligand and the ten secondary mushroom metabolites were loaded for grid preparation. Exhaustiveness for the docking was set at 8. Coordinates of the grid were set as 82.33 * 72.75 * 25.

Chapter 4

Result and Discussion

1. Virtual screening of the drugs against Nsp15 pre-docking results

After the docking of the 6vww macromolecule with the three drug ligands the results were as shown in table 1. The ligand of Nirmatrelvir was noticed to have the least binding energy at -7.6 with a root mean square deviation (RMSD) of 0.0. The three drugs were then superimposed with the co-crystal of 6vww to complete experimental and docking pose steps. The concluded root mean square deviation RMSD of all three of the drugs was seen to be <=2.0 Å as shown in the table 2.

Molecules	**Binding energy**
min_6vww_nirmatrelvir_uff_E=1796.64	-8.3
min_6vww_ritonavir_3D_uff_E=944.99	-8.2
min_6vww_remdesivir_uff_E=1212.27	-7.3

Table 1 - Docking results of the drugs with the nsp15 macromolecule taken from 6vww protein molecule.

Drugs Ligands	**RMSD**
Nirmatrelvir	1.15
Ritonavir	1.83
Remdesivir	1.89

Table 2 - RMSD calculated for experimental and docking pose step by superimposing co-crystal of 6vww protein with the drugs.

Selection of top lead from drugs and nsp15 docking

Since the ligand of Nirmatrelvir was observed with the least amount of binding energy at -8.3 kcal/mol, the compound was chosen as the top lead for further procedure in nsp15 antiviral discovery.

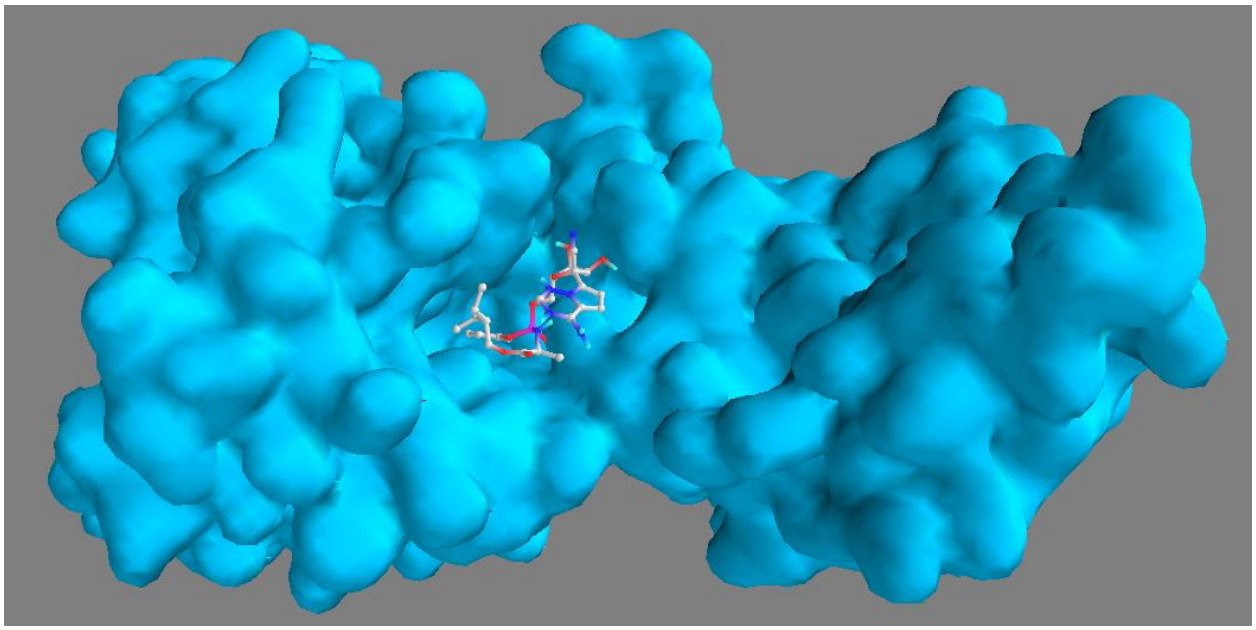


Fig19. Nirmatrelvir docked with Nsp15

2. Virtual Screening of mushroom metabolites against Nsp15

After the affirmation of the experimental and docking pose from the pre-docking results, the virtual screening of the secondary mushroom metabolites was performed. The ligand dataset of 10 compounds was subjected to virtual screening against Nsp15 protein target of COVID-19. The virtual screening was performed usingautodock tool implemented in PyRx.

The results of the virtual screening were obtained in a tabular form as shown in table 4. Compounds of Erinacine-A (PubChem CID: 10410568) and Ganodermadiol (PubChem CID: 139586903) were noticed to have binding energy less than that of the selected drug Nirmatrelvir, which was observed with a binding energy of -8.3 kcal/mol. The ligand of Erincine-A was observed with a binding energy of -8.4 kcal/mol while, the ligand of Ganodermadiol was observed with a binding energy of -8 kcal/mol.

Mushroom Metabolites Ligands	Binding energy
min_6vww_10410568_uff_E=723.17	-8.4
min_6vww_139586903_uff_E=792.58	-8
min_6vww_10789991_uff_E=793.63	-8
min_6vww_182264_uff_E=820.93	-7.9
min_6vww_17747381_uff_E=577.02	-7.5
min_6vww_42608175_uff_E=331.38	-7
min_6vww_101316898_uff_E=491.87	-6.7
min_6vww_101153516_uff_E=645.31	-6.6
min_6vww_102117112_uff_E=567.60	-6.3
min_6vww_10901419_uff_E=322.22	-6.2

Table 3 - Docking results of the secondary mushroom metabolites with the Nsp15 macromolecule taken from 6vww protein molecule.

Selection of top lead from mushroom metabolites.

Since the ligand of Erinacine-A was observed with the least amount of binding energy, the compound was chosen as the top lead for nsp15 antiviral discovery.

3. Molecular interaction analysis of top leads from Virtual screening of drugs and mushroom metabolites with Nsp15.

The Molecular interactions analysis were performed. This was analyzed through the Biovia Discovery Studio software. Ligands were loaded and selected and their 2D diagrams were generated. For hydrogen bond interaction, conventional hydrogen bonds were visualized and compared as seen in the results given in the figures below.

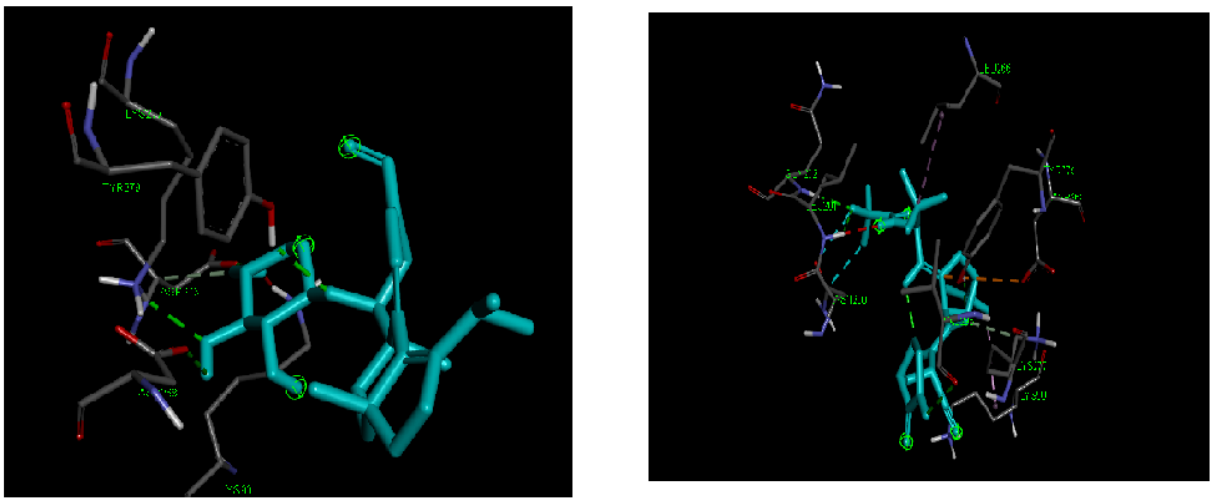


Fig20. 3D structure of molecular interactions of Nirmatrelvir and Erinacine-A

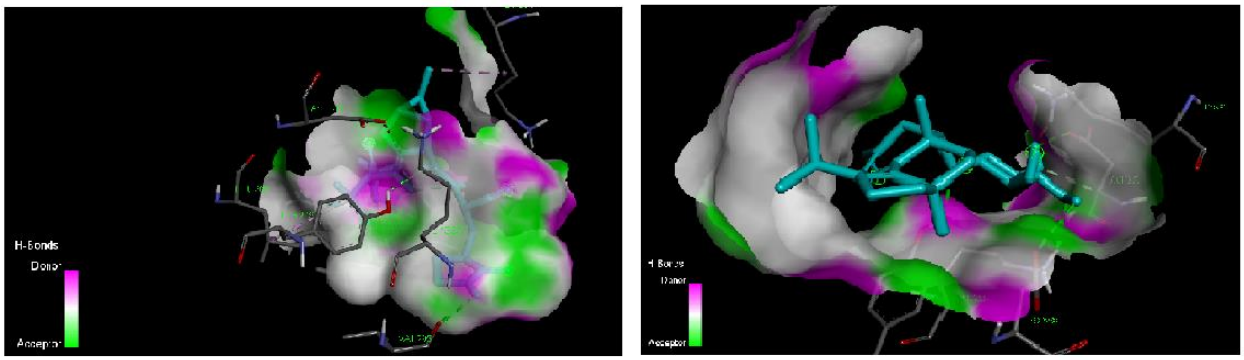


Fig21. 3D struture of H-bond interaction of Nirmatrelvir and Erinacine-A

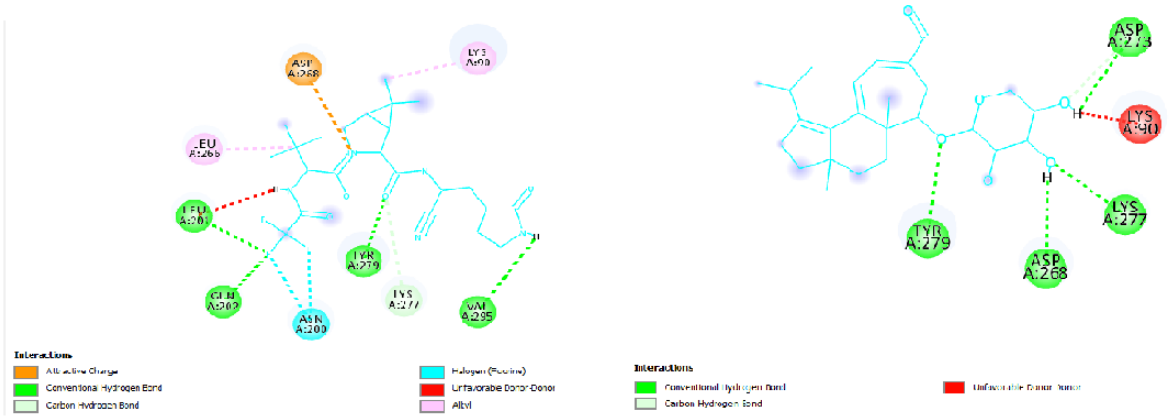


Fig22. 2D structure of Nirmatrelvir and Erinacine-A

Chapter – 5

Conclusion

Conclusion

Through the research and experiments conducted above, it can be concluded that the secondary mushroom metabolite Erinacine-A (PubChem CID: 10410568) that is known to have anti-viral properties and is found a promising therapeutic target against SARS-CoV-2 targeting specially Nsp15. This could be concluded because Erinacine-A was observed to have the least binding energy (-8.4) amongst the 10 secondary mushroom metabolites that were docked with the Nsp15 macromolecule.

References

1. Wang, D, et al. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus – infected pneumonia in Wuhan, China, 1061-1069 (2020).

2. Perlman, S. & Netland, J Coronavirus post – SARS: update on replication and patjogenesis. *Nat. Rev. Microbiol.* 7, 439-450 (2009)

3. G. Mariano, R. J. Farthing, S. L. M. Lale-Farjat, and J. R. C. Bergeron, "Structural Characterization of SARS-CoV-2: Where We Are, and Where We Need to Be," *Frontiers in Molecular Biosciences*, vol.7, p.605236, 2020, doi: 10.3389/fmolb.2020.605236.

4. Singh, A, K., Singh, A., Shaikh, A., Singh, R. & Misra, A. Chloroquine and hydroxychloroquine in the treatment of COVID-19 with or without diabetes: a systematic search and a narrative review with a special reference to India and other developing countries. *Diabetes Metab. Syndr*. 14, 241-246 (2020).

5. Thomas, Sunil. "Mapping the Nonstructural Transmembrane Proteins of Severe Acute Respiratory Syndrome Coronavirus 2." *Journal of Computational Biology*, 28 June 2021, 10.1089/cmb.2020.0627.

6. R. Raj, "Analysis of non-structural proteins, NSPs of SARS-CoV-2 as targets for computational drug designing," *Biochemistry and Biophysics Reports*, vol. 25, p. 100847, Mar. 2021, doi: 10.1016/j.bbrep.2020.100847.

7. Al-Rashedi, Nihad A. M., et al. "Prediction of Potential Inhibitors against SARS-CoV-2 Endoribonuclease: RNA Immunity Sensing." *Journal of Biomolecular Structure and Dynamics*, 27 Dec. 2020, pp. 1–14, 10.1080/07391102.2020.1863265. Accessed 6 May 2021.

8. Cornillez-Ty C.T., Liao L., Yates J.R., Kuhn P., Buchmeier M.J. Severe acute respiratory syndrome coronavirus nonstructural protein 2 interacts with a host protein complex

involved in mitochondrial biogenesis and intracellular signaling. *J. Virol.* 2009;83:10314–10318. doi: 10.1128/JVI.00842-09.

9. Lei, Jian, et al. “Nsp3 of Coronaviruses: Structures and Functions of a Large Multi-Domain Protein.” *Antiviral Research*, vol. 149, 1 Jan. 2018, pp. 58–74, www.ncbi.nlm.nih.gov/pmc/articles/PMC7113668/, 10.1016/j.antiviral.2017.11.001.

10. Fung TS, Liu DX, et al. “Post-translational modifications of coronavirus proteins: roles and functions.” *Future Virol* 13, 405-430 (2018).

11. Antonio Bensussen, Elena R. Alveraz-Buylla, Jose Diaz, et al. “SARS-CoV-2 Nsp5 Protein causes Acute Lung Inflammation, A dynamical Mathematical Model.” *Front Syst Biol*, 3 Dec. 2021

12. Peng Q, Peng R, Yuan B, Zhao J, Wang M, Wang X, Wang Q, Sun Y, Fan Z, Qi J, Gao GF, Shi Y, et al. “Structural and Biochemical Characterization of nsp12-nsp7-nsp8 Core Polymerase Complex from SARS-CoV-2.” *Cell Rep*, Jun 16. 2020

13. de O Araújo, Jéssica, et al. “Structural, Energetic and Lipophilic Analysis of SARS-CoV-2 Non-Structural Protein 9 (NSP9).” *Scientific Reports*, vol. 11, no. 1, 26 Nov. 2021, p. 23003, pubmed.ncbi.nlm.nih.gov/34837010/, 10.1038/s41598-021-02366-0. Accessed 1 May 2022.

14. Geoff Sutton, Elizabeth Fry, Lester Carter, Sarah Sainsbury, Tom Walter, Joanne Nettleship, Nick Berrow, Ray Owens, Robert Gilburt, Andrew Davidson, Stuart Siddell, Leo L M Poon, Jonathan Diprose, David Alderton, Martin Walsh, Jonathan M Grimes, David | Stuart, et al. “The Nsp9 Replicase Protein of SARS-coronavirus, structural and functional insights.” 12 Feb. 2004, doi: 10.1016/j.str.2004.01.016

15. Krafcikova, Petra, et al. “Structural Analysis of the SARS-CoV-2 Methyltransferase Complex Involved in RNA Cap Creation Bound to Sinefungin.” *Nature Communications*, vol. 11, no. 1, 24 July 2020, p. 3717, www.nature.com/articles/s41467-020-17495-9, 10.1038/s41467-020-17495-9. Accessed 12 Oct. 2021.

16. Mickael Bouvet, Adrien Lugari, Clara C Posthuma, Jessika C Zevenhoven, Stephanie Bernard, Stephane Betzi, Isabelle Imbert, Bruno Canard, Jean-Claude Guillemot, Patrick Lecine, Susanne Pfefferle, Christian Drosten, Eric J Snijder, Etienne Decroly, Xavier Morelli, et al. “Coronavirus Nsp10, a critical co-factor for activation of multiple replicative enzymes.” *J Biol Chem*, 12 Sept. 2014, doi: 10.1074/jbc.M114.577353.

17. Kundlik Gadhave, Prateek Kumar, Ankur Kumar, Taniya Bhardwaj, Neha Garg, Rajanish Giri, et al. “Nsp11 of SARS-CoV-2 is an Intrinsically Disordered Protein.” *Microbial Pathogenesis*, 7 Oct. 2020, doi: 10.1101/2020.10.07.330068

18. William haseltine, et al. “Functions of SARS-CoV-2 protein nsp12 in Polymerase Replication Transcription and Suppression of Natural Immunity.” *Heathcare Forbes*, 27 Jan. 2022,

19. Ma, Yuanyuan, et al. “Structural Basis and Functional Analysis of the SARS Coronavirus Nsp14–Nsp10 Complex.” *Proceedings of the National Academy of Sciences*, vol. 112, no. 30, 9 July 2015, pp. 9436–9441, 10.1073/pnas.1508686112.

20. Youngchang Kim, Robert Jedrzejczak, Natalia I Maltseva, Mateusz Wilamowski, Michae Endres, Adam Godzik, Karolina Michalska, Andrzej Joachimiak, et al.“Crystal Structure of Nsp15 Endoribonuclease NendoU from SARS-CoV -2.” *Protein Science*, 2 May 2020, 10.1002/pro.3873.

21. Monica C. Pillan, Meredith N. Frazier, Lucas B. Dillard, Jason G. Williams, Seda Kocaman, Juno M. Krahn, Lalith Perera, Cassandra K. Hayne, Jacob Gordon, Zachary D. Stewart, Mack Sobhany, Leesa J. Deterding, Allen L. Hsu, Venkata P. Dandey, Mario J. Borgnia, Robin E. Stanley, et al. "Cryo-EM structures of the SARS-CoV-2 endoribonuclease Nsp15 reveal insight to nuclease specificity and dynamics." *Nature Communications,* 27 Jan. 2021, doi: 10.1038/s41467-020-20608-z

22. Barage, Sagar, et al. "Identification and Characterization of Novel RdRp and Nsp15 Inhibitors for SARS-COV2 Using Computational Approach." *Journal of Biomolecular Structure and Dynamics,* 6 Nov. 2020, pp. 1–18, 10.1080/07391102.2020.1841026.

23. PubChem, "Nirmatrelvir," *pubchem.ncbi.nlm.nih.gov,* 1 May. 2021, https://pubchem.ncbi.nlm.nih.gov/compound/155903259

24. PubChem, "Remdesivir," *pubchem.ncbi.nlm.nih.gov,* 6 Aug, 2016, https://pubchem.ncbi.nlm.nih.gov/compound/121304016

25. PubChem, "Ritonavir" *pubchem.ncbi.nlm.nih.gov,* 26 Mar. 2005, https://pubchem.ncbi.nlm.nih.gov/compound/ritonavir