DEMOCRATIC PEOPLE'S REPUBLIC OF ALGERIA
MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH

Yahia Farès University of Médéa
FACULTY OF TECHNOLOGY
DEPARTMENT OF PROCESS ENGINEERING AND ENVIRONMENT

Doctoral Thesis

In: Process Engineering

Major: Pharmaceutical Engineering

Topic

Study and Development of QSAR Techniques for the Discovery of Therapeutically Relevant Molecules: Prediction of Potential Inhibitors of SARS-CoV-2

By Achouak Madani

Before the jury composed of:

Othmane Benkortbi	Thesis Advisor	UYFM	Professor
Maamar Laidi	Thesis Co-Advisor	UYFM	Professor
Latifa Khaouane	President	UYFM	Professor
Samira Ferhat	Examiner	UYFM	Professor
Houcine Boutoumi	Examiner	USDB 1	Professor
Salah Hanini	Guest	UYFM	Professor

Dedication

This thesis stands as a tribute to the steadfast support of my family. To my beloved mother, father, sister, and two brothers, your enduring encouragement and sacrifices have been the bedrock of my academic journey.

To my mother, whose constant support, love, sacrifice, and guidance have shaped my aspirations and given me strength during challenging times.

To my father, your unwavering beliefs in my abilities and your consistent encouragement have been pivotal in my pursuit of knowledge.

To my sister, your steadfast presence and understanding have provided solace and motivation throughout this endeavor.

To my two brothers, your persistent camaraderie and encouragement have added joy and purpose to this academic pursuit.

To my family, your resolute faith in me has been an endless source of inspiration. I am profoundly grateful for your unending support and love.

With heartfelt gratitude,

Achouak Madani

Acknowledgments

First and foremost, I thank Allah, the Almighty, the Merciful, who, by His grace, has led me to this stage in life.

I would like to express my deep appreciation to my thesis supervisor, Mr. Othmane Benkortbi, a professor at the University of Médéa, for his significant scientific concern demonstrated in the completion of this work.

I extend my thanks to my co-supervisor, Mr. Maamar Laidi, a professor and Vice Chancellor at the University of Médéa, for his involvement, assistance, guidance, and supervision in the completion of this thesis project.

Special thanks and gratitude go to Pr. Cherif Si-Moussa for his availability, help, and valuable input and time.

I also extend my sincere gratitude to Dr. Hentabli Hamza, researcher at the Chemistry Department, University of Missouri, USA , for his immense help in providing and facilitating internship opportunities.

I express my deepest gratitude to Pr. Naomie Salim, a Professor and the head of the Big Data Center at Universiti Teknologi Malaysia UTM, for supervising me during my internship at UTM and for her help, time, and effort. I thank also Dr. Haslina Hashim, a lecturer at the Faculty of Computing at UTM, and Dr. Hasmerya Marouf, a lecturer at the Faculty of Chemistryat UTM, for their precious help, time, and advice.

I express my gratitude to Pr. Latifa Khaouane, lecturer at the University of Médéa, for the honor of chairing my thesis jury.

Special thanks go to Pr. Houcine Boutoumi, lecturer at the University of Saad Dahlab Blida 1, for agreeing to review this work and participate in my thesis jury.

My sincere thanks go to Pr. Samira Ferhat, lecturer and head of the Process Engineering department, Faculty of Technology at the University of Médéa, for agreeing to judge this work.

I would also extend my deepest thanks and appreciations to Pr. Salah Hanini for his help and precious pieces of advice and for accepting the invitation to join the Jury panel as a guest.

I also wish to thank Mr. Bouaziz Mohamed Nadjib, Professor at the University of Médéa and director of the Laboratory of Biomaterials and Transport Phenomena (LBMPT), for his availability.

I also want to express my gratitude to all the teachers in the Department of Process Engineering and Environment for their contribution to the success of my education.

Abstract

The purpose of this work was to suggest two-classes inhibitors of 3CLpro enzyme for SARS-CoV-2 inhibitors based on structure: ADMET properties and molecular docking using ligand-based drug design approaches : QSAR: PSO-SVR and DA-SVR. The modelling of ML300 inhibitors was achieved through the development of a robust PSO-SVR model (R^2=0.96 and RMSE=0.13) that outperformed MLR and SVR models. The study suggested that branching degree 6 had a strong negative impact, while the moment of inertia X/Z ratio, the fraction of rotatable bonds, autocorrelation ATSm2, Keirshape2, and weighted path of length 2 positively impacted the activity. These outcomes proved the robustness of the PSO-SVR model and pave the way for its prediction abilities for future screening of more significant inhibitors' datasets.

The modelling of heterogeneous inhibitors through DA-SVR model (R^2=0.92and RMSE=0.21) that outperformed the BR-BPNN and DA-CNN models, that was endorsed by ADMET properties and molecular docking studies revealed that molecules 19 (-7.8 kcal/mole) and 15 (-7.5 kcal/mol) were the most affine to the binding site among 24 selected drug-like molecules with 1 kcal/mol difference less than the reference moleculePF-07321332, with conventional Hydrogen bonds HBs, pi-donor HBs, halogen, and alkyl bonds. They share triazole and 1H-isoindole fragments with fluorine in the 1H-isondole in molecule 15. The results of this study pave the way for QSAR-based synthesis and analog generation with respect to molecules 19 and 15 and the structural features indicated and predict their IC_{50} using the QSAR model to identify de novo potential inhibitors that can undergo wet-lab experiments.

Keywords: SARS-CoV-2; 3CLpro; QSAR; ADMET; Molecular docking.

ملخص

هذا العمل يقترح مثبطات من فئتين مختلفتين لإنزيم 3CLpro لفيروس كورونا باستخدام الطرق التصميمية للأدوية التي تعتمد على المركب العلاجي(العلاقة الكمية بين البنية والنشاط: PSO-SVR) والتي تعتمد على الهيكل (خصائص ADMETوالمحاكاة الجزيئية).

تم نمذجة مثبطات ML300 من خلال تطوير نموذج PSO-SVR قوي (R^2=0.96) وRMSE=0.13) الذي تفوق على نماذج MLR وSVR . تشير الدراسة إلى أن درجة التفرع 6 كان لها تأثير سلبي قوي، بينما كان لنسبة العزم X/Z، وكسر الروابط الدورانية، والتباينATSm2 ، Keirshape2، والمسار المرجح بطول 2 تأثير إيجابي على النشاط. تثبت هذه النتائج أن نموذج PSO-SVR قوي ودقيق، ويمهد الطريق لقدراته التنبؤية لفحص المستقبل لمجموعات بيانات لمثبطات أكثر أهمية .

وفيما يتعلق بنمذجة المثبطات الغير متجانسة من خلال نموذج DA-SVR (R^2=0.92, RMSE=0.21) الذي تفوق على نماذج BR-BPNN, DA-CNN والمدعوم بخصائص ADMET ودراسات المحاكاة الجزيئية، أظهر أن الجزيئات 19(-7.8 كيلو حريرة/مول) و 15 (-7.5 كيلو حريرة/مول) كانت الأكثر جاذبية للمكان الرابط بين 24 جزيئًة مختارة مناسبة للدواء بفارق 1 كيلو حريرة/مول أقل من جزيء المرجع PF-07321332 ، مع روابط هيدروجين تقليدية، وروابط pi-donor ، وروابط هالوجين، وروابط الألكيل. تشترك هذه الجزيئات في الأجزاء المشتركة لثنائي الترازول و 1 H-isoindole مع الفلور في 1 H-isoindole في الجزيء 15. تمهد نتائج هذه الدراسة الطريق لتخليق مركبات جديدة باستخدام تقنيات QSAR وإنشاء تناظريات جزيئية بالنسبة للجزيئات 19 و 15 والميزات الهيكلية المشيرة إليها، ويمكن توقع IC_{50} باستخدام نموذج QSAR لتحديد مثبطات جديدة يمكنها خضوع لتجارب المختبرات.

الكلمات المفتاحية: كوفيد-19، 3-سي ال برو، QSAR، خصائص ال ADMET، المحاكاة الجزيئية

Résumé

Le but de ce travail était de suggérer des inhibiteurs en deux classes de l'enzyme 3CLpro pour les inhibiteurs du SARS-CoV-2 basés sur la structure : les propriétés ADMET et le docking moléculaire en utilisant des approches de conception de médicaments basées sur les ligands : QSAR : PSO-SVR et DA-SVR. La modélisation des inhibiteurs ML300 a été réalisée grâce au développement d'un modèle robuste PSO-SVR (R^2=0.96 et RMSE=0.13) qui a surpassé les modèles MLR et SVR. L'étude a suggéré que le degré de ramification 6 avait un impact négatif fort, tandis que le ratio d'inertie X/Z, la fraction de liaisons rotatives, l'autocorrélation ATSm2, Keirshape2 et le chemin pondéré de longueur 2 ont eu un impact positif sur l'activité. Ces résultats ont prouvé la robustesse du modèle PSO-SVR et ouvrent la voie à ses capacités de prédiction pour le dépistage futur de ensembles d'inhibiteurs plus significatifs.

La modélisation des inhibiteurs hétérogènes via le modèle DA-SVR (R^2=0.92 et RMSE=0.21) qui a surpassé les modèles BR-BPNN et DA-CNN, soutenu par les propriétés ADMET et les études de docking moléculaire, a révélé que les molécules 19 (-7.8 kcal/mol) et 15 (-7.5 kcal/mol) étaient les plus affines au site de liaison parmi 24 molécules sélectionnées ressemblant à des médicaments, avec une différence de 1 kcal/mol par rapport à la molécule de référence PF-07321332, avec des liaisons hydrogène conventionnelles, des liaisons donneuses pi, halogène et alkyle. Elles partagent des fragments de triazole et de 1H-isoindole avec du fluor dans le 1H-isoindole de la molécule 15. Les résultats de cette étude ouvrent la voie à la synthèse basée sur le QSAR et à la génération d'analogues pour les molécules 19 et 15, en tenant compte des caractéristiques structurales indiquées, et prédisent leur IC50 en utilisant le modèle QSAR pour identifier de nouveaux inhibiteurs potentiels pouvant faire l'objet d'expériences en laboratoire humide.

Mots clé : SARS-CoV-2 ; 3CLpro ; QSAR ; ADMET ; Docking moléculaire

List of Abbreviations

Chapter I: SARS-CoV-2

ACE2 Angiotensin Converting Enzyme 2

ALT Alanine Aminotransferase

ARDS Acute Respiratory Distress Syndrome

ARDS Acute Respiratory Distress Syndrome

AST Aspartate Aminotransferase

COVID-19 Coronavirus Disease 2019

CRP C Reactive Protein

CT Computed Tomography

ER Endoplasmic Reticulum

ERGIC Endoplasmic Reticulum Golgi Intermediate Compartment

ExoN Exoribonuclease

FDA Food and Drug Administration

GSK3 Glycogen Synthase Kinase 3

HIV-1 Human Immunodeficiency Virus 1

IFIT1 Interferon Induced Protein with Tetratricopeptide Repeats 1

IFITMs Interferon Stimulated Gene Products

IL Interleukin

IL-6 Interleukin 6

LDH Lactate Dehydrogenase

LY6E Lymphocyte Antigen 6 Family Member E

MERS-CoV Middle East Respiratory Syndrome Coronavirus

N7-MTase N7 Methyltransferase

NF-κB Nuclear Factor kappa B

NSP Non-structural Protein

ORF Open Reading Frame

PDB-ID Protein Data Bank Identification

PDE5 Phosphodiesterase 5

PHEIC Public Health Emergency of International Concern

PRRA Polybasic Cleavage Site

pp1ab Polyprotein 1ab

RBD Receptor Binding Domain

RBM Receptor Binding Motif

RdRp RNA dependent RNA Polymerase

RNA Ribonucleic Acid

RTC Replication and Transcription Complex

S1/S2 Spike protein subunits 1 and 2

SAM S-adenosyl L methionine

SARS-CoV-2 Severe Acute Respiratory Syndrome Coronavirus 2

TMPRSS2 Transmembrane Protease Serine 2

TLRs Toll like Receptors

Chapter II: Computer-Aided Drug Design

ACO Ant Colony Optimization

ADMET Absorption Distribution Metabolism Excretion and Toxicity

AI Artificial Intelligence

ANN Artificial Neural Networks

BBB Blood Brain Barrier

CNN Convolutional Neural Networks

CYP Cytochrome P450

CoMFA Comparative Molecular Field Analysis

CoMSIA Comparative Molecular Similarity Indices Analysis

DNN Deep Neural Networks

DA Dragonfly Algorithm

ECFP Extended Connectivity Fingerprints

GA Genetic Algorithm

GWO Grey Wolf Optimizer

HIA Human Gastrointestinal Absorption

HSG Hydrogen Suppressed Graph

IC_{50} Half Maximal Inhibitory Concentration

IIC Index of Ideality Correlation

kNN k Nearest Neighbors

LBDD Ligand Based Drug Design

LogP Partition Coefficient

MACCS Keys Molecular ACCess System Keys

MD Molecular Dynamics

MC Monte Carlo

MLR Multiple Linear Regression

NMR Nuclear Magnetic Resonance

NB Naive Bayesian

P-gp P glycoprotein

PDB Protein Data Bank

PCR Principal Component Regression

PCA Principal Component Analysis

PLS Partial Least Squares

PSO Particle Swarm Optimization

QSAR Quantitative Structure Activity Relationship

RBN Restricted Boltzmann Machines

RF Random Forest

RNN Recurrent Neural Networks

SA Simulated Annealing

SBDD Structure Based Drug Design

SAR Structure Activity Relationships

SMILES Simplified Molecular Input Line Entry System

SVM Support Vector Machine

VS Virtual Screening

WOA Whale Optimization Algorithm

X-ray X ray Crystallography

Chapter III: Modelling procedure

AD Applicability Domain

AIC Akaike's Information Criterion

BR-BPNN Bayesian Regulation Backpropagation Neural Network

DA-CNN Dragonfly Algorithm Convolutional Neural Network

DCV Double Cross Validation

Da Dalton (unit of mass)

MMF94 Merck Molecular Force Field 94

MLPCN Molecular Libraries Probe Production Centers Network

PSO-SVR Particle Swarm Optimization Support Vector Regression

RMSE Root Mean Square Error

SDF Structure Data File

SD Standard Deviation

Chapter IV: Results and Discussions

RBF Radial Basis Function

VIF Variance Inflation Factor

List of Figures

Chapter I SARS-CoV-2

Fig I.1 Timeline of the key events the pandemic 6

Fig I.2 Death cases distribution region wise from December 2019 to December 2022 7

Fig I.3 Reported cases (a) and deaths (b) in Algeria from February2020 to December 2022 on a weekly basis 9

Fig I.4 COVID19 symptoms 10

Fig I.5 SARS-CoV-2 structure (a) schematic diagram, and (b) microscopic image 12

Fig I.6 Comparison of coronavirus spike proteins (a) gene sequence of the spike protein of SARS-CoV-2 and other CoVs, and (b) structural comparison 15

Fig I.7 Entry steps of SARS-CoV-2 to host cells 20

Fig I.8 Cleavage of the Spike protein by Furin 21

Fig I.9 membrane fusion induced by the SARS-CoV-2 S protein 21

Fig I.10. Overview of the SARS-CoV-2 Life Cycle 23

Fig I.11 Distribution of various SARS-CoV-2 clades and lineages from Algeria 25

Fig I.12 Vaccination coverage in the world (WHO) 27

Chapter II Computer-Aided Drug Design

Fig II.1 SBDD workflow 37

Chapter III Modelling Procedure

Fig III.1 organizational chart of the ML300 analogues modelling 61

Fig.III.2 2D structures of noncovalent and covalent inhibitors: (a) ML300,(b) GC376 62

Fig III.3 organizational chart of the heterogeneous inhibitors modelling 69

Fig III.4 Samples from the 246 selected inhibitors 70

Chapter IV Results and Discussions

Fig IV.1 The relative importance of the selected descriptors 79

Fig IV.2 MLR regression plot 80

Fig IV.3 SVR Regression plot 81

Fig IV.4 PSO-SVR Regression plot 82

Fig IV.5 Applicability domain of (a) MLR model, (b) SVR model, and (c) PSO-SVR model 85

Fig IV.6 Pairwise correlation plot of the selected descriptors 92

Fig IV.7. Descriptors' relative importance (a) word cloud, (b) histogram 94

Fig IV.8 BR-BPNN regression plot 95

Fig IV.9 Architecture of the DA-CNN model 97

IV.10 DA-CNN regression plot 98

Fig IV.10 DA-SVR regression plot 106

Fig IV.11 Applicability domain of the DA-SVR model 108

Fig IV.12 R^2 and RMSE bar plot for some literature-reported models and the present DA-SVR model 109

Fig IV.13 Structure of molecules with low GIA (molecule 6) and violations of the Pfizer Rule 110

Fig IV.14 Interactions of molecules (a) 15, (b): 19, (c): 7, (d): 13, (e): 18, (f): 24 115

Fig IV.15 Interactions of PF-07321332 (a) best pose, (b): second best pose 116

List of Tables

Chapter II Computer-Aided Drug Design

Table II.1 Docking softwares 38

Table II.2 Tools for pharmacophore modeling 47

Table II.3 Molecular descriptors and fingerprints calculation tools 49

Chapter III Modelling Procedure

Table III.1 ML300 analogues database 63

Table III.2 Validation criteria 68

Table III.3 Major ADMET properties for the selected 24 molecules 72

Chapter IV Results and Discussions

Table IV.1 Statistical criteria for the MLR, SVR, and PSO-SVR models 80

Table IV.2 The experimental and predicted values of pIC_{50} predicted by the PSO-SVR model, train values in pink, and test values in green 82

Table IV.3 External validation criteria for the MLR, SVR, and PSO-SVR models 84

Table IV.4 Comparison between literature and the present paper 88

Table IV.5 The selected molecular descriptors 91

Table IV.6 VIF and tolerance values for the selected descriptors 91

Table IV.7 Statistical validation criteria for the BR-BPNN model 94

Table IV.8 The statistical validation criteria for the DA-CNN model 97

Table IV.9 The structure of the DA-SVR model 99

Table IV.10 DA-SVR experimental and predicted values, train values in pink, and test values in green 99

Table IV.11 Statistical validation criteria for the DA-SVR model 106

Table IV.12 Major rules evaluation of the 24 molecules 111

Table of content

List of Abbreviations

List of Figures

List of Table

General Introduction 2

Chapter I: SARS-CoV-2

1. Introduction 6
2. Emergence and spread 6
3. COVID-19 in Algeria 8
4. Symptoms 10
5. Diagnosis 11
6. Genomics and phylogeny 12
7. Transmission 16
8. Infection cycle 17
   - 8.1. Entry mechanism 17
   - 8.2. Viral Protein Translation Process 22
   - 8.3. Variants of concern in Algeria 24
   - 8.4. BA.2.86 variant (Pirola): A new VOC? 24
   - 8.5. What the future may hold 25
9. Vaccines 26
   - 9.1. Approved vaccines 27
     - 11.1.1. Nucleic acid vaccines: 27
     - 11.1.2. Non-replicating viral vectored vaccines: 27
     - 11.1.3. Inactivated virus vaccines: 27
     - 11.1.4. Recombinant subunit vaccines: 28
   - 9.2. Vaccines Comparison: Pfizer, Moderna, and Johnson & Johnson 28
10. Potential drugs 29
    - 10.1. Novel drugs in phase III 29
      - 1.1.1. PF-07321332(Paxlovid) 29
      - 1.1.2. S-217622 29
      - 1.1.3. AT-9010 30
    - 10.2. Drug repurposing 30
      - 1.1.4. RNA-Dependent RNA Polymerase Inhibitors 30

1.1.5. Ebselen 31
1.1.6. Arbidol 31
1.1.7. Interferons 31
1.1.8. Protease Inhibitor 31
1.1.9. Ivermectin 31
1.1.10. Melatonin 31
1.1.11. Teicoplanin 31
1.1.12. 5-Alpha Reductase Inhibitors 32
1.1.13. Spironolactone 32
1.1.14. Sildenafil/Tadalafil 32
1.1.15. Statins 32
1.1.16. N-Acetylcysteine 32
1.1.17. Bromhexine Hydrochloride 32
1.1.18. Chloroquine Phosphate or Hydroxychloroquine 32
1.1.19. Corticosteroids 32
10.3. Traditional medicine 32
11. Conclusion 33

Chapter II: Computer-Aided Drug Design

1. Introduction 35
2. Computer-aided drug design 35
2. Structure-based drug design 36
2.1. Molecular docking 37
2.2. Molecular docking procedure 39
2.2.1. Macromolecule preparation 39
2.2.2. Small molecules preparation 40
2.3. Molecular Dynamic MD Simulation 40
2.4. ADMET properties 41
2.4.1. Absorption 41
2.4.2. Distribution 41
2.4.3. Metabolism 41
2.4.4. Excretion 41
2.4.5. Toxicity 41
3. SBDD literature review 42
4. Ligand-based drug design 46
4.1. Pharmacophore modeling 47

4.2. Quantitative Structure-Activity Relationship QSAR 48
4.3. Artificial Intelligence 52
5. LBDD literature review 53
6. Virtual screening 56
6.1. Structure-Based Virtual Screening (SBVS) 57
6.2. Ligand-Based Virtual Screening (LBVS) 57
7. Conclusion 57

Chapter III: Modelling Procedure

1. Objectives 60
2. Methodology 60
2.1. ML300 analogues 60
2.1.1. Data set retrieval 60
2.1.2. Descriptors generation and selection 65
2.1.3. Data division and model development 65
2.1.4. Regression methods 66
2.1.5. Statistical validation 66
2.1.6. Applicability domain 66
2.2. Heterogeneous inhibitors 67
2.2.1. Dataset retrieval and pretreatment 67
2.2.2. Molecular descriptors calculation and selection 67
2.2.3. Database division 67
2.2.4. Regression methods 70
2.2.5. Validation metrics 71
2.2.6. Applicability domain 71
2.2.7. ADMET properties 71
2.2.8. Molecular docking 71

Chapter IV: Results and Discussion

1. ML300 analogues 76
1.1. Selection of representative descriptors 76
1.1.1. Constitutional descriptors 76
1.1.2. Topological descriptors 76
1.1.3. Geometrical descriptors 77
1.2. QSAR models' development 78
1.2.1. MLR model 78
1.2.2. Support vector regression model 79

1.2.3. SVR-particle swarm optimization model 80
1.2.4. External validation 83
1.2.5. Applicability domain 83
1.2.6. Models' comparison 85
1.2.7. Comparison with literature 86
2. Heterogeneous datasets: 88
2.1. Descriptors' mechanistic interpretation 88
2.2. BR-BPNN model 93
2.3. DA-CNN model 95
2.4. DA-SVR model 98
2.5. Models' comparison 106
2.6. Applicability domain 107
2.7. Comparison with literature 107
2.8. ADMET properties 108
2.9. Molecular docking 111
3. Conclusion 115
General Conclusion 119
References 121

Introduction

General Introduction

In the vast landscape of contemporary science, certain challenges emerge as crucibles of innovation, pushing the boundaries of human knowledge and demanding ingenious solutions. Among these, the COVID-19 pandemic, caused by the novel coronavirus SARS-CoV-2, constitutes an unprecedented global crisis. The pandemic's devastating impact on public health, economies, and societies at large has created an urgent need for effective therapeutics. In the fight against this formidable foe, computational chemistry has emerged as a beacon of hope, leveraging the power of advanced algorithms to unravel the mysteries of molecular interactions and accelerate drug discovery processes.

At the heart of this research lies the profound intersection of computational chemistry and the fight against SARS-CoV-2. In the face of a virus that has reshaped the world in unprecedented ways, this study embarks on a nuanced exploration into the intricate world of Computer-Aided Drug Design techniques. The central question of this thesis echoes across laboratories and research institutions worldwide: how can computational methodologies, specifically CADD techniques, contribute to the prediction and development of inhibitors capable of disrupting the virulence mechanisms of SARS-CoV-2?

SARS-CoV-2, a member of the coronavirus family, has a particular molecular architecture, enabling its virulent entry and replication within human cells. The virus's intricate components, particularly the 3CLpro enzyme, serve as promising targets for therapeutic intervention. Understanding the subtle nuances of the interaction between potential inhibitors and these components is critical for the design of effective drugs. However, the sheer complexity of these interactions requires sophisticated computational approaches, marking the genesis of this research endeavor.

In the arsenal of modern drug discovery, computational chemistry is emerging as a powerful force, bridging the chasm between theoretical insights and practical applications. It embodies the fusion of mathematics, physics, and biology, encapsulated within intricate algorithms. In this field, QSAR techniques stand out as indispensable tools, offering a systematic framework to correlate the chemical structures of molecules with their biological activities. In the context of SARS-CoV-2, QSAR techniques provide a unique perspective through which the inhibitory potential of compounds can be dissected, paving the way for targeted drug design.

The journey embarked upon in this thesis is guided by a multifaceted methodological approach. At its core, the study harnesses the innovative amalgamation of Particle Swarm Optimization (PSO) with Support Vector Machines (SVM) and Differential Evolution Algorithm (DA) with SVM. These techniques, often regarded as the vanguard of computational prediction, possess the ability to decipher complex patterns in large data datasets. The synergy between PSO and SVM, as well as DA and SVM, represents not only a computational strategy but also an intellectual endeavor to push the boundaries of predictive modeling.

Recognizing the multifaceted nature of drug development, this research extends its gaze beyond QSAR techniques. The study meticulously delves into the realm of Absorption, Distribution, Metabolism, Excretion, and Toxicity (ADMET) properties, highlighting the importance of a holistic approach to drug design. ADMET properties serve as stringent filters, ensuring that potential inhibitors not only exhibit strong interactions with the target but also possess the required bioavailability and safety profiles, essential for their transformation into effective therapeutics.

Simultaneously, the research unfurls the intricate art of molecular docking, a computational simulation technique that illuminates the potential binding affinities and interaction patterns between inhibitors and viral components. The three-dimensional exploration of these interactions not only informs the design process but also offers valuable information on the inhibitory potential of the modeled compounds. It is within these virtual landscapes that the theoretical hypotheses of inhibitory efficacy are rigorously tested and refined.

It is essential to note that this study does not adhere to a one-size-fits-all approach. Rather, it embraces the complexity of the challenge by modeling two distinct categories of inhibitors. The first category includes analogues of ML300, a known inhibitor, requiring detailed exploration to improve their efficacy. These analogues, while rooted in the existing body of knowledge, demand a nuanced understanding of their unique structural intricacies to optimize their inhibitory potential.

This research extends its purview to heterogeneous inhibitors. This category represents the diverse array of potential compounds that could serve as inhibitory agents against SARS-CoV-2. It encompasses myriad chemical structures, each with its own fingerprint of inhibitory potential. Exploring these heterogeneous compounds reflects the complexity of real-world drug discovery, where diversity is often the key to new therapeutic solutions.

This thesis, spanning the domains of computational chemistry, biology, and pharmacology, is not merely an academic endeavor; it is a testament to human ingenuity in the face of crisis. It signifies a synergy between scientific curiosity and societal need, illuminating a path toward innovative drug design. As the chapters unfold, they weave a narrative of a collective endeavor, where scientists and researchers collaborate across disciplines and continents, pooling their expertise to confront a global threat.

This thesis provides, in a first place, a deep explanation of SARS-CoV-2, than it dives in CADD and its tools with an extensive literature review in chapter II. Chapter III details the methodology used to predict the inhibitors and concerns both the ligand and structure based approaches. Chapter IV then presents the rich results found along with a profound interpretation of the various results of QSAR modelling and molecular docking. A general conclusion is presented at last to highlight the key points and findings of this thesis.

CHAPTER I
SARS-CoV-2

1. Introduction

The emergence of SARS-CoV-2, and thus the COVID-19 pandemic, in 2019 has caused millions of cases and deaths worldwide. Understanding the structure and mechanisms of infection of this novel virus has therefore become mandatory to fight it. This chapter covers the details of this pandemic, from emergence, genome analysis, and symptoms to recent variants of concern, vaccines and potential drugs.

2. Emergence and spread

More than 767.518 million cases and 6.974 million deaths have been reported since December 2019 (www.who.org), the first outbreak of the deadly Corona Virus Disease 2019 COVID-19 pandemic. The new causative virus was named Severe Acute Respiratory Syndrome Coronavirus 2 SARS-CoV-2 on 11 February 2020 by the International Committee on Taxonomy of Viruses ICTV. The key events are shown in Fig I.1. The World Health Organization WHO declared it as a Public Health Emergency of International Concern (PHEIC) on 30 January 2020, but did not qualify the outbreak as a pandemic until 11 March 2020[1]. This delayed international response to the pandemic hindered the efficient containment of the pandemic as lockdown measurements and international mobility restrictions were not applied promptly Fatal consequences quickly followed both on the sanitary and economic axes. In addition to millions of deaths, the pandemic has caused the worst global economic crisis since the 1900s. The crisis increased levels of poverty and unemployment around the world and had a dramatic impact on inequality within and between countries. Small businesses, freelancers and women were among the most impacted due to the lockdown and increasing debts (www.worldbank.org).

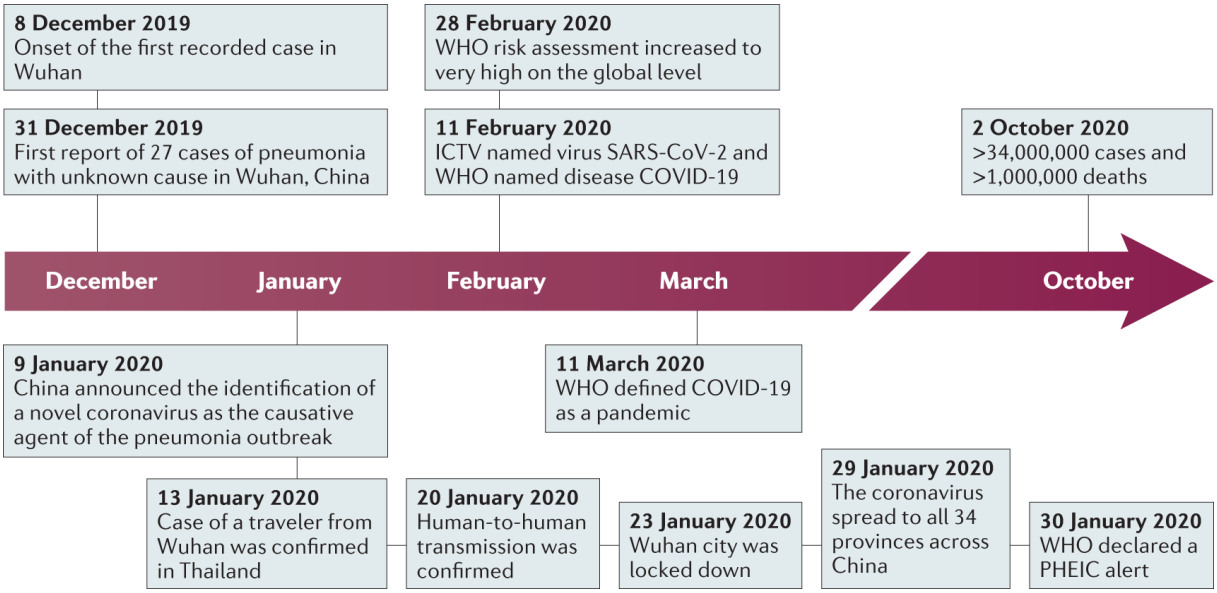


Fig I.1 Timeline of the key events the pandemic

The number of reported cases and deaths varies across regions and continents. America being the most influenced continent with 2 957 229 deaths, followed by Europe with 2 243 918 deaths. Africa, according to the statistics, recorded the lowest death cases with 175399 official deaths, as shown in Fig I.2. This is due to a combination of factors, including the relatively lower population density, younger demographic profile, and timely implementation of preventive measures. Africa, compared to other continents, faced certain advantages in containing the spread of the disease.

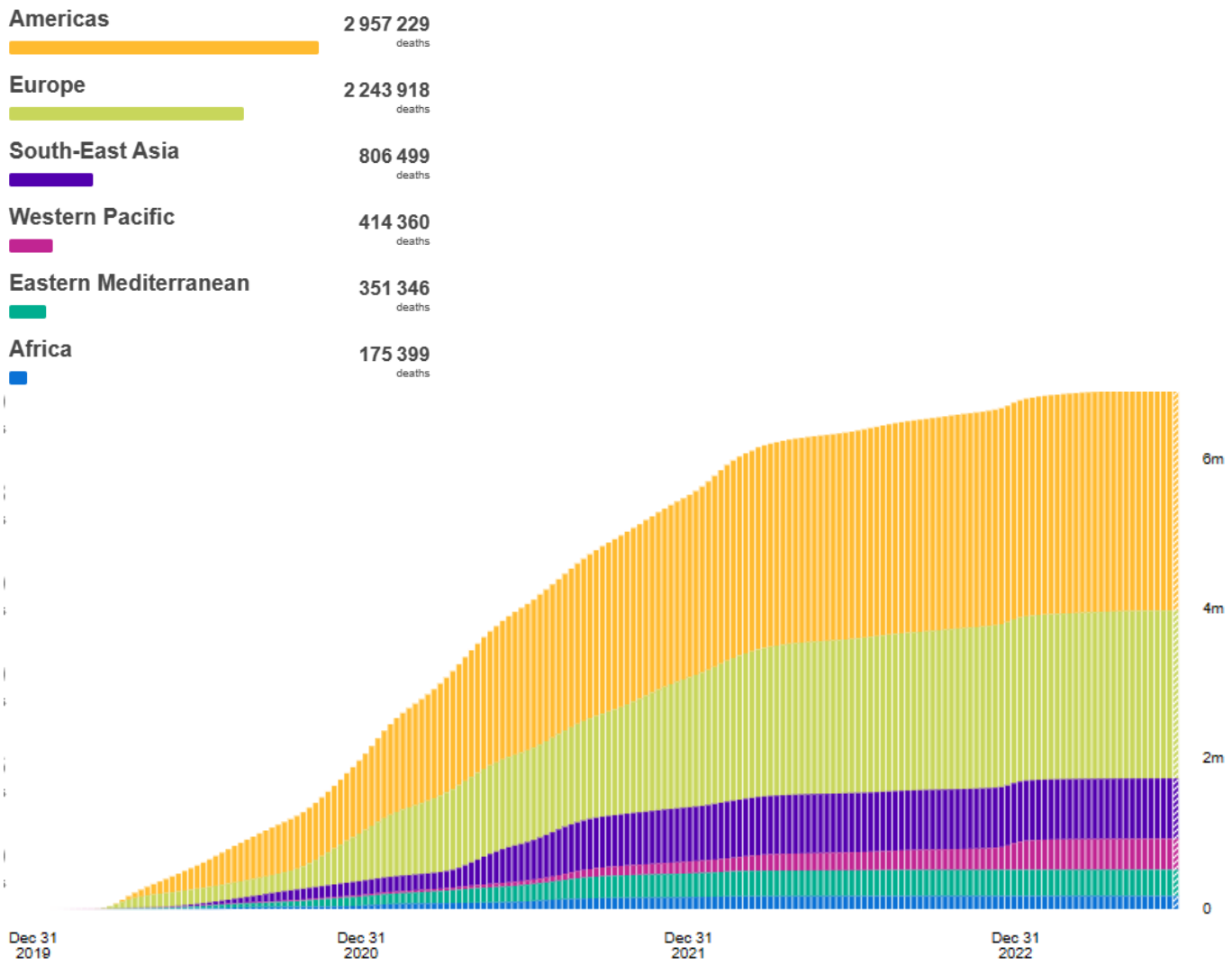


Fig I.2 Death cases distribution region wise from December 2019 to December 2022

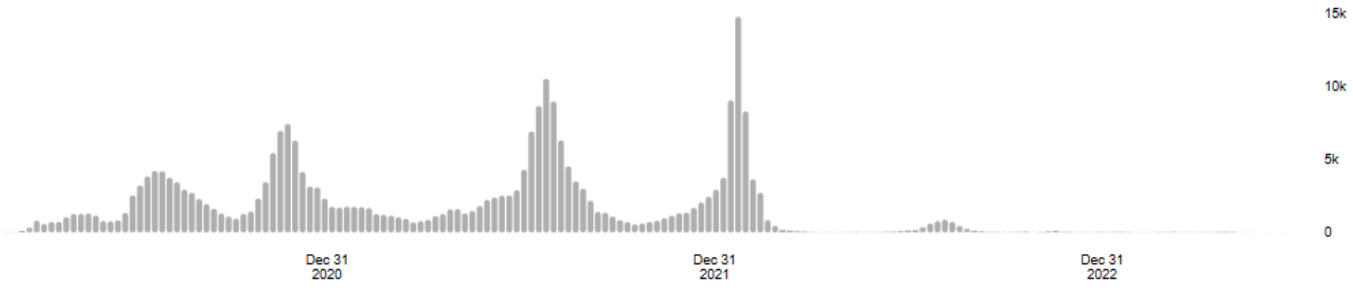
Population density is an important factor contributing to the decline in the number of deaths in Africa. Many African countries have a lower population density compared to regions like Europe and America. This reduced population density can help in limiting the transmission of the virus, as it becomes more challenging for the virus to spread rapidly in areas with fewer people in close proximity. Furthermore, Africa has a relatively younger population compared

to other continents. COVID-19 has shown to have a more severe impact on older individuals and those with underlying health conditions. The younger demographic in Africa might have played a role in mitigating the severity of the disease, as younger individuals generally have a higher likelihood of recovering from the virus compared to older age groups. Additionally, African countries took swift and proactive measures to combat the spread of COVID-19. Many nations implemented strict public health measures, including early lockdowns, travel restrictions, and widespread testing and contact tracing. These measures helped to contain the spread of the virus and prevent the healthcare systems from becoming overwhelmed. It's also important to consider that the reported number of deaths may not reflect the true extent of COVID-19-related mortality in Africa. Limited testing capacity, underreporting, and challenges in data collection and reporting systems could contribute to the lower number of official deaths.

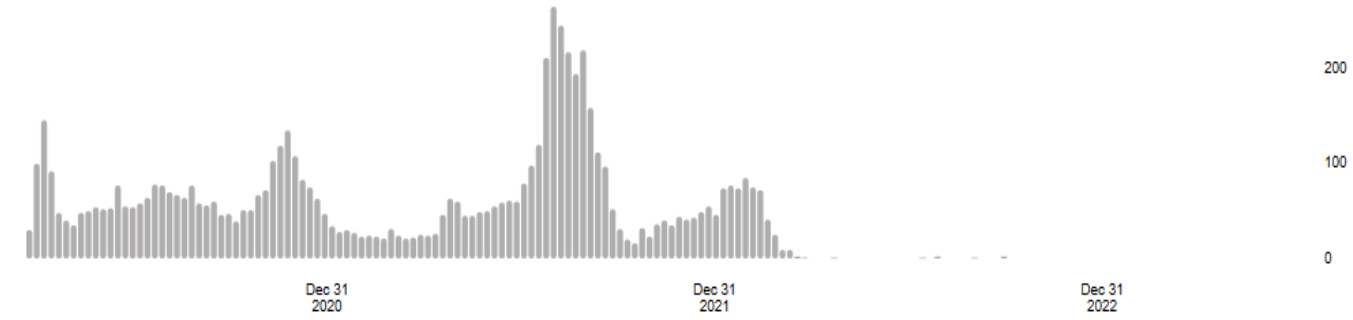
3. COVID-19 in Algeria

The COVID-19 pandemic entered to Algeria with the detection of the first confirmed case on February 25th, 2020[3]. Swiftly, the virus began to spread across the country, and the province of Blida became the epicenter of the Algerian epidemic[4]. The initial phase of the pandemic in Algeria presented a range of challenges, including a significant number of imported infections, limited capacity for COVID-19 testing, and shortcomings in surveillance efforts[3]. As the number of cases continued to rise, Algeria found itself among the three African countries deemed at highest risk for COVID-19 importation, along with Egypt and South Africa[5]. The situation intensified when, on March 1st, two new cases were reported in the Blida region. These cases were linked to Algerian nationals who had traveled from France. The emergence of this cluster marked the beginning of a challenging period for the region, which subsequently experienced a surge in infections. Over the following months, the number of cases steadily increased throughout the country[4]. This is illustrated in Fig I.3, as it shows the reported cases and deaths from the first recorded case to December 2022. Statistics from WHO report that Algeria recorded 271 852 confirmed cases, and 6 881 deaths. As shown in the figure, four peaks are distinguished: June-July 2020, November 2020, July-August 2021, and January 2022. These peaks are due to *(i)* relaxation of the containment measures, and *(ii)* emergence of novel variances of the virus. 26 January 2022 recorded the highest number of confirmed cases of 2521 and a daily increase of 306 cases, while 29 July 2021 was the deadliest with 49 deaths in 24 hours. Algeria faced notable hurdles in its response to the pandemic. The laboratory analysis and testing capacity were identified as weak links in the system, leading to a limited number of daily tests being conducted and longer waiting times for obtaining test results. This situation created a higher risk of asymptomatic individuals unknowingly spreading the virus before being identified and isolated[4,6]. To combat the spread of the virus, the Algerian government implemented a series of containment and quarantine measures. These measures, which became more stringent on March 24th, included the prohibition of sports gatherings, the closure of schools, places of worship, mosques, and the suspension of public transportation. In addition, travel restrictions were imposed, with flights and maritime connections to and from Algeria being halted[5,7]. In terms of treatment protocols, Algeria was among the early adopters of the

hydroxychloroquine/azithromycin protocol, following press releases from the World Health Organization (WHO). Initially, this protocol showed promising results, with a decrease in the death rate observed after its widespread implementation. However, as the situation evolved, the protocol was limited to severe cases and reserved as a last resort treatment option[4,8]. The social impact of the pandemic in Algeria was profound. A study analyzing population density and COVID-19 cases revealed that heavily populated cities like Algiers, Blida, Oran, and Setif were the most affected[9]. There was a stronger correlation between population density and case numbers in coastal regions compared to less densely populated highland areas. The relationship between population density and COVID-19 spread was less significant in the southern region of the country[9]. Daily reported cases in Algeria displayed fluctuations



(a)



(b)

Fig I.3 Reported cases (a) and deaths (b) in Algeria from February2020 to December 2022 on a weekly basis

over time. After peaking from late April to mid-May, a decrease was observed, attributed to various interventions including border closures, school shutdowns, social distancing measures, and the cancellation of public gatherings. However, a subsequent rise in daily reported cases occurred in mid-June 2020, partially linked to the relaxation of containment measures and a gradual return to normalcy[10]. The impact of the pandemic on education in Algeria was substantial. Teachers had to swiftly adapt their traditional teaching methods to online platforms, which presented numerous challenges. While distance learning emerged as a practical solution in urgent cases, the long-term effects on the educational system in Algeria remained uncertain[11–13]. Overall, the COVID-19 epidemic in Algeria brought to light a

range of challenges, from testing capacity issues and treatment protocols to social and educational impacts. The country witnessed fluctuations in daily reported cases and implemented a variety of measures to mitigate the spread of the virus.

4. Symptoms

Patients infected with SARS-CoV-2 can exhibit a broad range of symptoms, spanning from mild to severe, with a significant proportion of individuals remaining asymptomatic carriers[14], as shown in Fig I.4. According to a study involving the determination of COVID19 clinical features[15], the most commonly reported symptoms include fever 81.2%, cough 58.5%, fatigue: 38.5%, Dyspnea: 26.1%, and the presence of Sputum: 25.8%, which are consistent with respiratory tract viral infections. Chest X-rays of patients with pneumonia often reveal multiple mottling and ground-glass opacity, indicative of lung involvement[16,17]. In addition to respiratory symptoms, COVID-19 has been associated with gastrointestinal

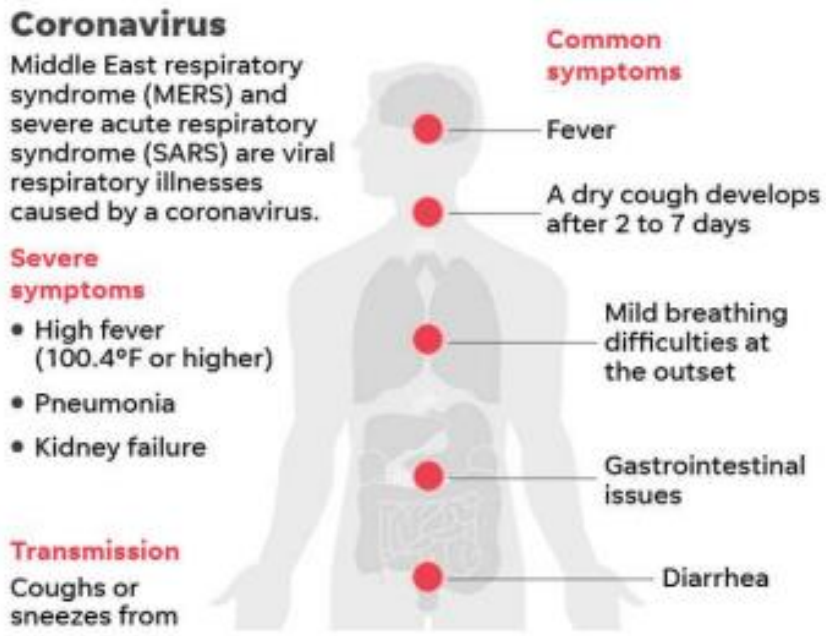


Fig I.4 COVID19 symptoms[18]

manifestations in a subset of patients. Gastrointestinal symptoms such as vomiting, diarrhea, and abdominal pain have been reported in approximately 2-10% of cases. Interestingly, in some instances, gastrointestinal symptoms like diarrhea and nausea have been observed to precede the development of fever and respiratory symptoms[17,19].

Upon laboratory examination, COVID-19 patients frequently exhibit alterations in their blood parameters. Decreased lymphocyte and eosinophil counts have been reported, while median hemoglobin values tend to be lower. Conversely, there are increases in white blood cell (WBC) and neutrophil counts, as well as elevated serum levels of C-reactive protein (CRP), lactate dehydrogenase (LDH), aspartate aminotransferase (AST), and alanine

aminotransferase (ALT)[20]. Notably, the initial CRP serum levels have emerged as a significant independent predictor for the progression of severe COVID-19 infection[21,22].

Although the primary target of coronavirus infection is the respiratory system, the widespread distribution of angiotensin-converting enzyme 2 (ACE2) receptors in various organs suggests the possibility of multi-organ involvement[23]. As a result, COVID-19 can potentially lead to cardiovascular, gastrointestinal, kidney, liver, central nervous system, and ocular damage[24]. Notably, the cardiovascular system is often affected in severe cases, with reported complications including myocardial injury, myocarditis, acute myocardial infarction, heart failure, dysrhythmias, and venous thromboembolic events. Monitoring with high sensitivity cardiac troponin may prove useful in such cases[25].

Severe cases of COVID-19 can rapidly progress to acute respiratory distress syndrome (ARDS), leading to multiple organ failure induced by the so-called "cytokine storm"[17]. The cytokine profile in severe COVID-19 cases resembles that of secondary hemophagocytic lymphohistiocytosis syndrome and is characterized by increased levels of interleukins (IL-2, IL-7), granulocyte colony-stimulating factor, interferon-c-inducible protein-10, monocyte chemoattractant protein 1, macrophage inflammatory protein 1- α, and tumor necrosis factor- α [26]. Moreover, elevated levels of ferritin and IL-6 have been identified as predictors of fatality, suggesting that hyper inflammation induced by the virus may contribute to severe outcomes[27].

In the context of children with COVID-19, the disease course is generally milder compared to that seen in adults, but severe and fatal cases have also been reported. Clinical laboratory data in pediatric cases exhibit differences from adults, with inconsistent alterations in the leukocyte index[28]. However, elevations in the levels of CRP, procalcitonin, and LDH have been observed in children with severe disease, raising concerns about potential cardiac involvement[29].

5. Diagnosis

Timely diagnosis plays a crucial role in containing the spread of COVID-19. Molecular detection of SARS-CoV-2 nucleic acid is considered the standard method[30]. Numerous viral nucleic acid detection kits targeting specific genes (ORF1b, N, E, or S) are commercially available with varying detection times, ranging from minutes to hours, based on the technology used[1].

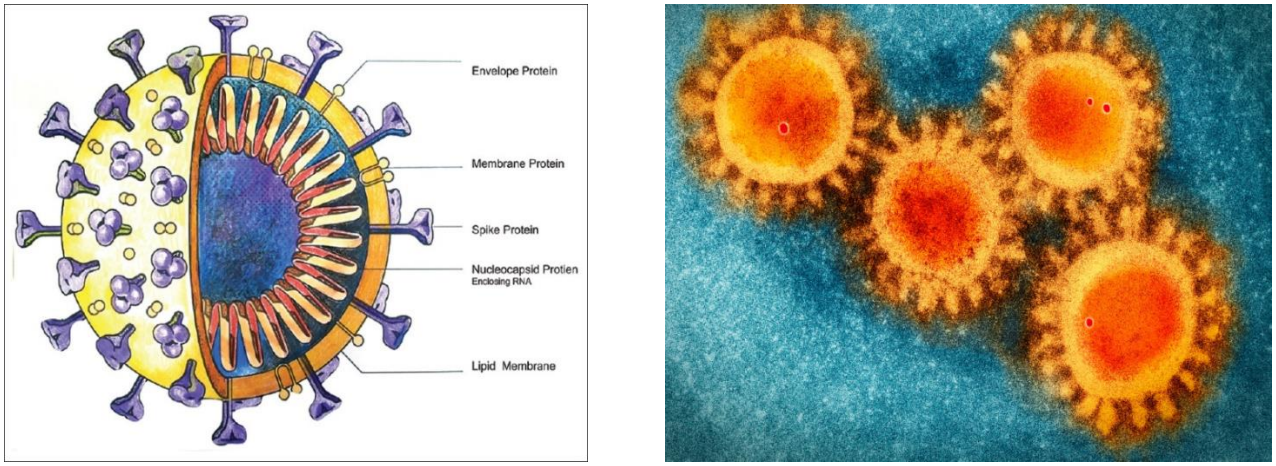
It's important to note that the accuracy of molecular detection can be influenced by various factors. While SARS-CoV-2 can be detected from different respiratory sources such as throat swabs, saliva, nasopharyngeal swabs, sputum, and bronchial fluid, lower respiratory tract samples typically exhibit higher viral loads[31–33]. Moreover, viral nucleic acid has been found in samples from the intestinal tract or blood, even when respiratory samples are negative[34]. Further, viral load may decline from its peak level upon disease onset, leading to potential false negatives when using oral swabs[33,35]. Henceforth, employing multiple detection methods is recommended to confirm a COVID-19 diagnosis.

In cases where molecular detection capacity is overwhelmed, chest Computed Tomography scans were utilized in Wuhan to quickly identify patients. Initial CT scans often revealed bilateral multilobar ground-glass opacities with a peripheral or posterior distribution, indicating potential COVID-19 infection[36,37]. Combining repeated swab tests with CT scans has been suggested for individuals showing high clinical suspicion of COVID-19 but testing negative in initial nucleic acid screenings[37].

To supplement molecular diagnosis, serological tests detecting antibodies to specific SARS-CoV-2 proteins (N or S) were employed, particularly in later stages of the disease or for retrospective studies[38]. However, the extent and duration of immune responses remain uncertain, and available serological tests vary in their sensitivity and specificity, necessitating careful consideration when selecting and interpreting their results. Additionally, potential future tests for T cell responses should be considered in further diagnostic advancements[1].

6. Genomics and phylogeny

SARS-CoV-2 belongs to the viral group of Coronaviruses, which are a large family of viruses that affect the lower respiratory tract, leading to a spectrum of illnesses ranging from mild cold-like symptoms to severe infections with a fatality rate of up to 50%[15]. Coronaviruses are named for the crown-like spikes that protrude from their surface, resembling the sun's corona[18], as shown in Fig I.5. They have a positive-sense single-stranded RNA genome and are known to infect a wide range of species, including bats, birds, and mammals. SARS-CoV-2 is not the first corona virus the world has encountered. There are



(a) *(b)*

Fig I.5 SARS-CoV-2 structure (a) schematic diagram, and (b) microscopic image[39]

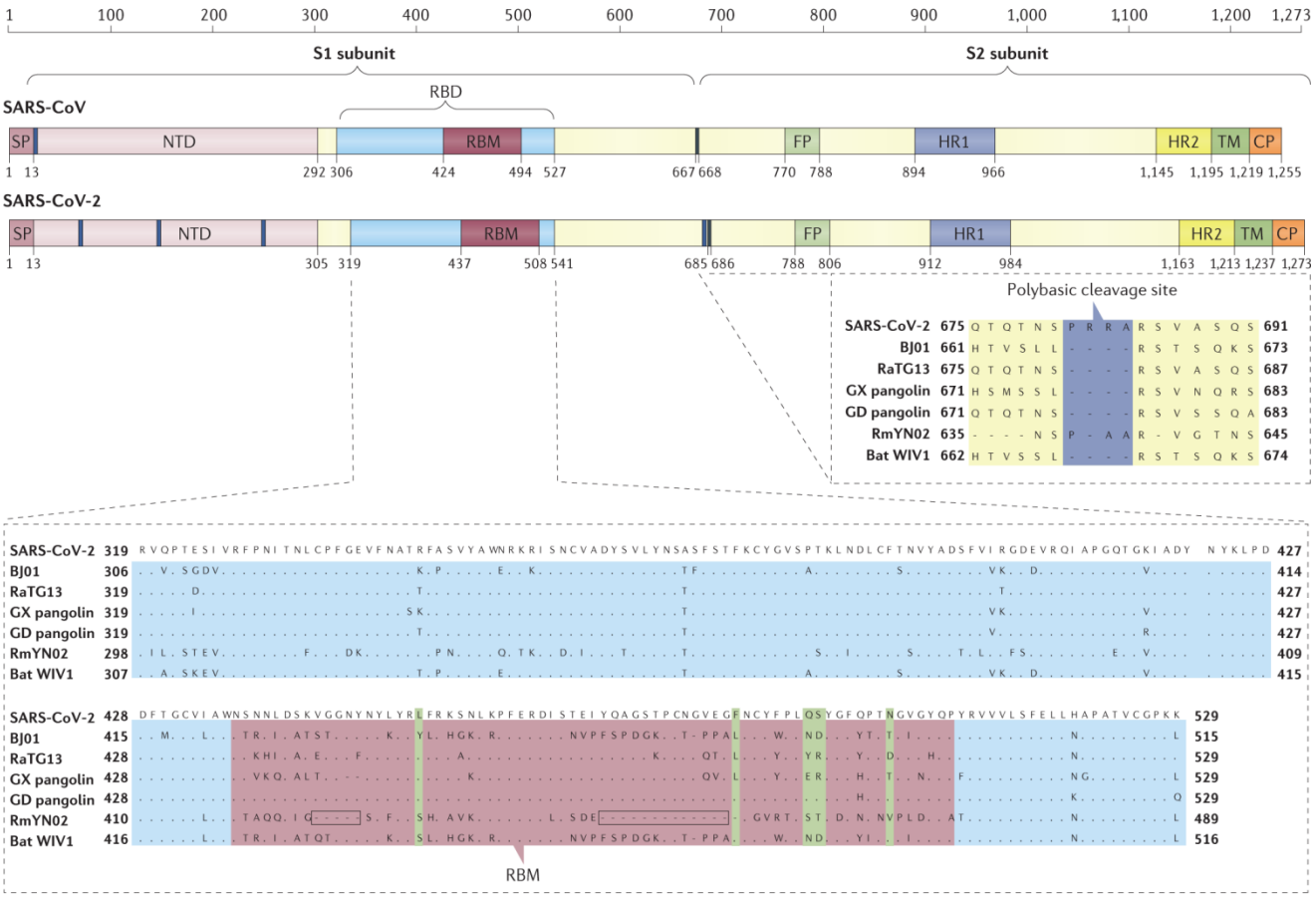
two other well-known Coronaviruses that have caused severe outbreaks in the past: SARS-CoV-1 and MERS-CoV[40]. SARS-CoV-1, which emerged in 2002, is the virus responsible for causing Severe Acute Respiratory Syndrome (SARS)[41]. It spread to more than two dozen countries and resulted in over 8,000 reported cases and approximately 800 deaths[42]. SARS-CoV-1 is believed to have originated in bats and was transmitted to humans through

intermediate hosts, such as civet cats. MERS-CoV, on the other hand, is the virus that causes Middle East Respiratory Syndrome (MERS)[41]. This virus was first identified in 2012 and primarily affects the respiratory system. MERS-CoV has been linked to camels as a possible intermediate host, and the majority of reported cases have been from countries in the Arabian Peninsula[43]. While the number of MERS cases has been relatively low compared to other Coronaviruses, the fatality rate associated with MERS is much higher, estimated at around 35%[1]. These outbreaks serve as reminders of the potential dangers posed by Coronaviruses and highlight the importance of surveillance, early detection, and rapid response to prevent the spread of these viruses[44]. It's worth noting that SARS-CoV-2, the virus responsible for the COVID-19 pandemic, shares some similarities with SARS-CoV-1 and MERS-CoV[45], but it is also distinct in many ways[46]. It is highly transmissible between humans, leading to a global health crisis[47].

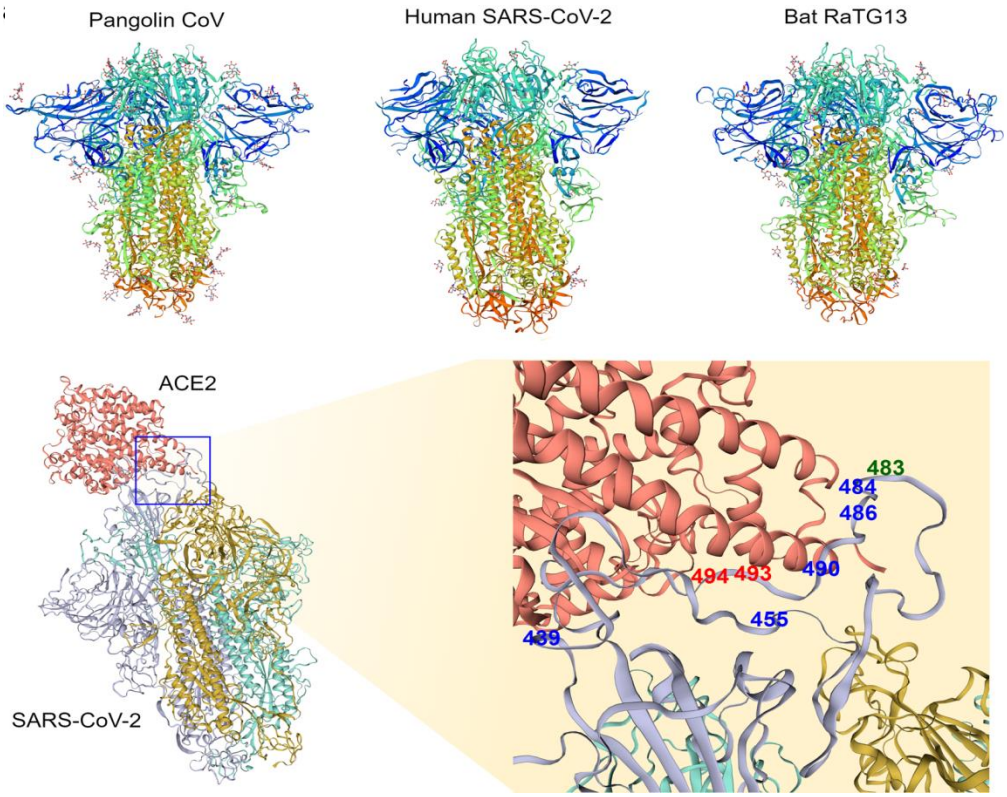
SARS-CoV-2 exhibits significant similarities of 79% with SARS-CoV-1 and 50% of MERS-CoV[48]. The genome organization of SARS-CoV-2 aligns with that of other betacoronaviruses, featuring six functional open reading frames (ORFs) arranged in a specific order from 5′ to 3′. These ORFs include replicase (ORF1a/ORF1b), spike (S), envelope (E), membrane (M), and nucleocapsid (N). Notably, interspersed between the structural genes are seven putative ORFs encoding accessory proteins[49]. Comparing the proteins encoded by SARS-CoV-2 to those of SARS-CoV, there is a striking similarity in terms of length, with most proteins being comparable in size. Among the four structural genes, SARS-CoV-2 shares over 90% amino acid identity with SARS-CoV, except for the S gene, which shows divergence[1]. The replicase gene, which constitutes a significant portion of the 5′ genome, encodes a large polyprotein known as pp1ab. Through proteolytic cleavage, this polyprotein generates 16 non-structural proteins responsible for vital functions like transcription and virus replication. Importantly, most of these non-structural proteins of SARS-CoV-2 display over 85% amino acid sequence identity with their counterparts in SARS-CoV[49].

In terms of the spike (S) protein, SARS-CoV-2 stands out with a full size of 1,273 amino acids, surpassing both SARS-CoV (1,255 amino acids) and known bat SARSr-CoVs (ranging from 1,245 to 1,269 amino acids). Fig I.6.a represents the comparison of the spike protein and other coronaviruses. Panel (a) shows a schematic diagram of the spike (S) protein of severe acute respiratory syndrome coronavirus (SARS-CoV) and SARS-CoV-2. It highlights insertions in the S protein, with an enlargement at the bottom right displaying the specific insertions at amino acids 675–691 in the SARS-CoV-2 S protein, aligned with those of other coronaviruses, including SARS-CoV BJ01, RaTG13, pangolin coronavirus reported from Guangdong, China (GD pangolin), pangolin coronavirus reported from Guangxi, China (GX pangolin), and bat SARS-related coronavirus (SARSr-CoV) WIV1. Panel (b) demonstrates the alignment of the receptor-binding domain (RBD) in these coronaviruses. The receptor-binding motif (RBM) is marked in purple, while the five crucial residues directly contacting angiotensin-converting enzyme 2 (ACE2) are highlighted in green. Fig I. 6.b represents the Structural comparison of coronavirus spike proteins. In the top half, it illustrates a long-axis trimer view of the closed conformation of cryo-EM spike protein structures from three coronaviruses: pangolin coronavirus (PDB ID: 7CN8, left panel), human

SARS-CoV-2 (PDB ID: 6ZB5, middle panel), and bat RaTG13 coronavirus (PDB ID: 7CN4, right panel). The models are color-coded in a rainbow pattern from the N-terminus (blue) to the C-terminus (red).In the bottom half, the left panel illustrates the human SARS-CoV-2 spike protein (colored by chain: purple, yellow, and green) in bound confirmation with the ACE2 complex (colored red) (PDB ID: 6ACG). The right panel zooms in on a specific region comprising 8 amino acids (positions shown in blue, green, and red) previously identified as targets of positive selection in various studies (references [50–52]). Positions in blue were



(a)



(b)

Fig I.6 Comparison of coronavirus spike proteins (a) gene sequence of the spike protein of SARS-CoV-2 and other CoVs[1]*, and (b) structural comparison*[50].

spositively selected in one of the cited studies, while positions in red (493 and 494) were identified in two studies. The green position 483 of the S protein was positively selected in all three studies. While the S protein of SARS-CoV-2 shares amino acid sequence similarities of 76.7-77.0% with SARS-CoVs from civets and humans, 75-97.7% with bat coronaviruses in the same subgenus, and 90.7-92.6% with pangolin coronaviruses, the amino acid similarity in the receptor-binding domain (RBD) between SARS-CoV-2 and SARS-CoV drops to only 73%[53,54]. Another distinct genomic feature of SARS-CoV-2 is the insertion of four amino acid residues (PRRA) at the junction of subunits S1 and S2 of the S protein[54]. The addition of this sequence results in the formation of a polybasic cleavage site (RRAR), which facilitates effective cleavage by furin and other proteases[55]. It is worth noting that this S1-S2 cleavage site is not observed in all related viruses within the subgenus Sarbecovirus, except for a similar three amino acid insertion (PAA) found in RmYN02, a bat-derived coronavirus recently identified in Rhinolophus malayanus in China[56]. While the insertion found in RmYN02 doesn't fully exhibit the properties of a polybasic cleavage site, it does offer evidence that this feature, previously believed to be exclusive to SARS-CoV-2, might have naturally developed[56]. According to structural studies, the furin-cleavage site can influence the stability of the SARS-CoV-2 S protein, assisting in the required conformational changes that enable the binding of the RBD (Receptor-Binding Domain) to its receptor[57]. Nevertheless, it remains to be proven whether the increased transmissibility of SARS-CoV-2

in comparison to SARS-CoV is directly linked to the acquisition of the furin-like cleavage site[54].

Furthermore, SARS-CoV-2 possesses an accessory gene called orf8, which encodes a novel protein showing only 40% amino acid identity to the ORF8 protein of SARS-CoV. Unlike the ORF8 protein of SARS-CoV, the one encoded by SARS-CoV-2 does not contain a motif that triggers intracellular stress pathways[49]. Significantly, there has been a detection of a SARS-CoV-2 variant in Singapore, featuring a 382-nucleotide deletion encompassing the entire ORF8. This deletion bears resemblance to the 29- or 415-nucleotide deletions observed in human SARS-CoV variants during the latter stages of the 2002-2003 outbreaks[58]. The deletions observed in ORF8 indicate the possibility of human adaptation after cross-species transmission[1].

To evaluate the genetic variation among different SARS-CoV-2 strains, researchers aligned 77,801 genome sequences from SARS-CoV-2 detected worldwide. Through this analysis, they identified a total of 15,018 mutations, including 14,824 single-nucleotide polymorphisms[59]. Within the spike protein, four amino acid alterations (V483A, L455I, F456V, and G476S) are situated near the binding interface of the RBD. However, the effects of these alterations on binding to the host receptor are currently unknown. Among the S variant sites, the alteration D614G in the S1 subunit has been observed more frequently than others and has become a significant marker for a major subclade of SARS-CoV-2 known as clade G[1,60]. Since March 2020, SARS-CoV-2 variants with the G614 alteration in the S protein have replaced the original D614 variants and become the dominant form globally. In patients infected with the G614 variant, higher viral loads have been detected, but clinical data does not suggest a significant association between the D614G alteration and disease severity[60]. Pseudo typed viruses carrying the S protein with the G614 alteration have exhibited higher infectious titers compared to those carrying the S protein with D614, indicating a potential increase in the infectivity of SARS-CoV-2[60]. However, it is essential to note that results obtained from in vitro experiments using pseudo-virus models may not precisely reflect natural infection. Therefore, further studies using wild-type SARS-CoV-2 variants and different target cells and animal models are needed to validate these findings[1]. Another noteworthy mutation in SARS-CoV-2 evolution is a single-nucleotide polymorphism at nucleotide position 28,144, resulting in an amino acid substitution of Ser for Lys at residue 84 of the ORF8 protein. Variants carrying this mutation form a distinct subclade referred to as "clade S"[50,61]. However, the available sequence data is currently insufficient to fully understand the early global transmission history of the virus. Factors such as travel patterns, founder effects, and public health measures play a significant role in the spread of specific lineages, regardless of potential biological differences among different virus variants[1].

7. Transmission

SARS-CoV-2 spreads efficiently through various transmission routes, with respiratory transmission being the primary mode of infection. Respiratory droplets released when an infected person coughs, sneezes, talks, or breathes can be inhaled by nearby individuals, leading to infection. Additionally, aerosols may play a significant role in transmission,

especially in crowded and poorly ventilated spaces[62,63]. Apart from respiratory transmission, emerging evidence suggests the possibility of the oral-fecal route as another means of virus dissemination. SARS-CoV-2 RNA has been detected in stool samples of patients with COVID-19[64], raising concerns about the potential role of sewage in viral transmission. Researchers are exploring biosorbents capable of retaining and inactivating the virus to mitigate any environmental transmission risks[65]. Furthermore, SARS-CoV-2 has been detected in saliva[66], attributed to the presence of ACE2 receptors in the salivary gland ducts. This finding adds to our understanding of how the virus may spread and highlights the importance of saliva-based testing for early detection and control of the disease[67]. Interestingly, urine samples from COVID-19 patients have shown viral RNA positivity. While the extent of viral shedding in urine and its potential infectivity remain to be established, this finding underscores the need for comprehensive testing methods to accurately diagnose COVID-19[68]. The virus can also survive on various surfaces, such as door handles and cell phones, posing a risk of infection to individuals who touch contaminated surfaces and subsequently touch their eyes, nose, or mouth. Ensuring regular disinfection of frequently touched surfaces and practicing good hand hygiene are essential preventive measures[62]. Regarding vertical transmission, the evidence remains inconclusive. Although there have been reports of no mother-to-child transmission during pregnancy or breastfeeding[69], a recent case involving an infant with elevated IgM against SARS-CoV-2 indicates the possibility of in utero infection[70]. Further research is required to fully understand the mechanisms of vertical transmission.

Additionally, the eyes may serve as a potential route of transmission. SARS-CoV-2 RNA has been detected in ocular swabs, suggesting the virus's presence in ocular secretions[71]. Inspite of the fact that data is lacking, taking precautionary measures such as wearing goggles while examining suspected or confirmed COVID-19 patients is advisable[72].

In summary, understanding the various transmission routes of SARS-CoV-2 is crucial for implementing effective public health strategies to control the spread of COVID-19. Combining rigorous preventive measures, widespread testing, and continued research will be essential in mitigating the impact of this ongoing global health crisis.

8. Infection cycle

8.1. Entry mechanism

The mechanism by which SARS-CoV-2 gains entry into human host cells involves a series of complex and finely orchestrated steps. The viral entry proteins play a critical role in this process, as they must first attain an energetically stable conformation, and subsequently undergo a conformational transition that provides the necessary energy to overcome the inherent repulsion between the virus and the cellular membranes. To achieve this delicate balance, the spike (S) protein of SARS-CoV-2 adopts a metastable state, which exhibits a tendency to transform into a lower-energy state prior to the actual membrane fusion event. This transition is facilitated by two proteolytic cleavage events that occur following the binding of the S protein to ACE2 receptors on the surface of host cells[73]. Fig I.7 illustrates

the entry steps. It shows the two cleavage sites that are crucial for entry: one at the S1-S2 junction during virus maturation (internal to S2 subunit) and the other at the S2′ site following ACE2 binding. Depending on the entry route, different proteases cleave the S2′ site[73].

Interestingly, the cleavage of the S protein at specific sites is a crucial prerequisite for the subsequent fusion process. The first cleavage site, known as the S1-S2 boundary, is localized to the junction between the S1 and S2 subunits of the spike protein. The second cleavage site, referred to as the S2' site, is found within the S2 subunit[74]. In the case of SARS-CoV, both of these sites are cleaved by proteases present in the target cell. However, for SARS-CoV-2, there is a difference in the cleavage pattern. The S1-S2 boundary is cleaved by the furin protease in the virus-producer cell (as shown in Fig I.8), while the cleavage at the S2' site still relies on proteases in the target cell[75,76]. The dependence on target-cell proteases for both viruses implies that the availability and activity of these proteases significantly impact the efficiency of viral entry[77].

Two primary proteases involved in the activation of the S protein are Transmembrane Protease Serine 2 TMPRSS2 and cathepsin L. TMPRSS2 is a serine protease with type II transmembrane characteristics[73], and although its specific physiological role and substrate specificity remain incompletely understood, it has undergone extensive research in relation to respiratory virus infections, including those caused by SARS coronaviruses. TMPRSS2 is expressed in various tissues, such as the gastrointestinal, respiratory, and urogenital epithelium, making it a key player in viral entry[78,79]. In the case of SARS-CoV-2, there are two possibilities for the viral entry: TMPRSS2-mediated S protein activation that occurs at the plasma membrane since this protease is present on the cell surface. On the other hand, there is the cathepsin-mediated activation that occurs within the endolysosomes. These two entry processes are fully illustrated in Fig 7. The presence of TMPRSS2 at the cell surface enables S protein activation to occur directly at the plasma membrane, while cathepsin-mediated activation relies on internalization of the virus-ACE2 complex into the endosomes for subsequent cleavage[73].

Once the S protein has undergone the necessary cleavage events, it undergoes a series of structural rearrangements to facilitate membrane fusion as shown in Fig I.9. The receptor-binding domain (RBD) of the spike protein plays a crucial role in this process. The RBD samples both "up" and "down" conformations due to intrinsic protein dynamics. However, binding of ACE2 captures the RBD in the "up" conformation, leading to the expulsion of the 630 loop and the fusion peptide proximal region (FPPR) from their positions in the closed S trimer conformation. This event exposes the S2' cleavage site near the fusion peptide for proteolytic cleavage. The dissociation of the 630 loop from the hydrophobic surface of the C-terminal domain 2 (CTD2) destabilizes the domain, freeing the N-terminal segment of S2 from S1. This results in the release of the S1 subunit altogether, facilitated by the precleavage of the S1-S2 boundary by furin protease. The dissociation of S1 initiates a cascade of refolding events in the S2 subunit, enabling the transition to a stable postfusion structure. In this process, the fusion peptide (FP) of S2 is inserted into the target cell membrane, and the heptad repeat 1 (HR1) undergoes unfolding[73,80,81], while HR2 folds back, bringing the

fusion peptide and transmembrane segments closer together. This rearrangement leads to the bending of the membranes, eventually facilitating membrane fusion[82].

Apart from the direct involvement of proteases and viral entry proteins, various cellular proteins also play a significant role in either restricting or promoting viral entry. Toll-like receptors (TLRs) are an essential component of the innate immune system and recognize pathogen-associated molecular patterns. When activated by the presence of SARS-CoV-2 RNA, certain TLRs trigger the production of proinflammatory cytokines, contributing to the host's immune response. Interferon-stimulated gene products (IFITMs) and lymphocyte antigen 6 family member E (LY6E) are cellular proteins that influence viral entry. IFITMs are induced by type I and type II interferons and restrict viral entry by preventing viruses from traversing the endosomal membrane to access the cellular cytoplasm. LY6E, a cell surface protein, has been found to inhibit S protein-mediated membrane fusion, potentially affecting the efficiency of viral entry.

The understanding of the intricate and multifaceted SARS-CoV-2 entry mechanism provides valuable insights that can aid in the development of antiviral strategies. By targeting specific steps involved in viral entry, researchers may identify potential therapeutic interventions to combat COVID-19 effectively. Continuous exploration and investigation into the molecular details of the virus-host interactions will be crucial in the ongoing battle against the global pandemic.

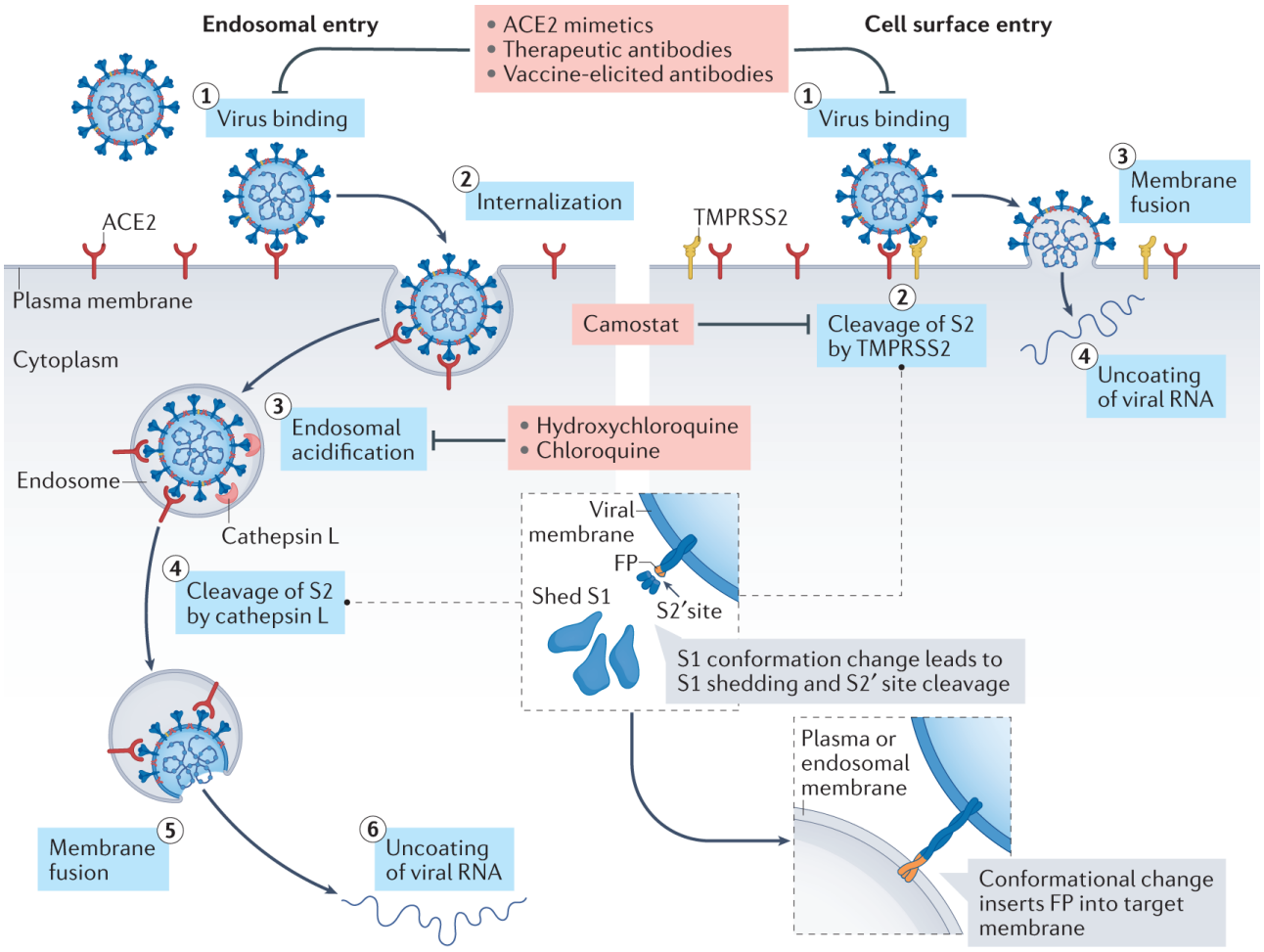


Fig I.7 Entry steps of SARS-CoV-2 to host cells[73]

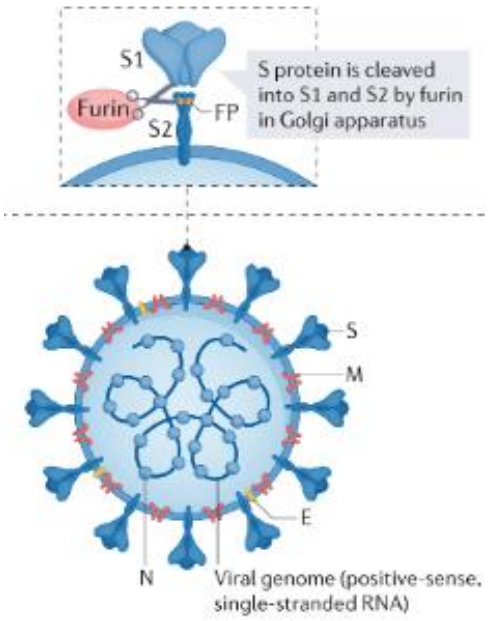


Fig I.8 Cleavage of the Spike protein by Furin[73]

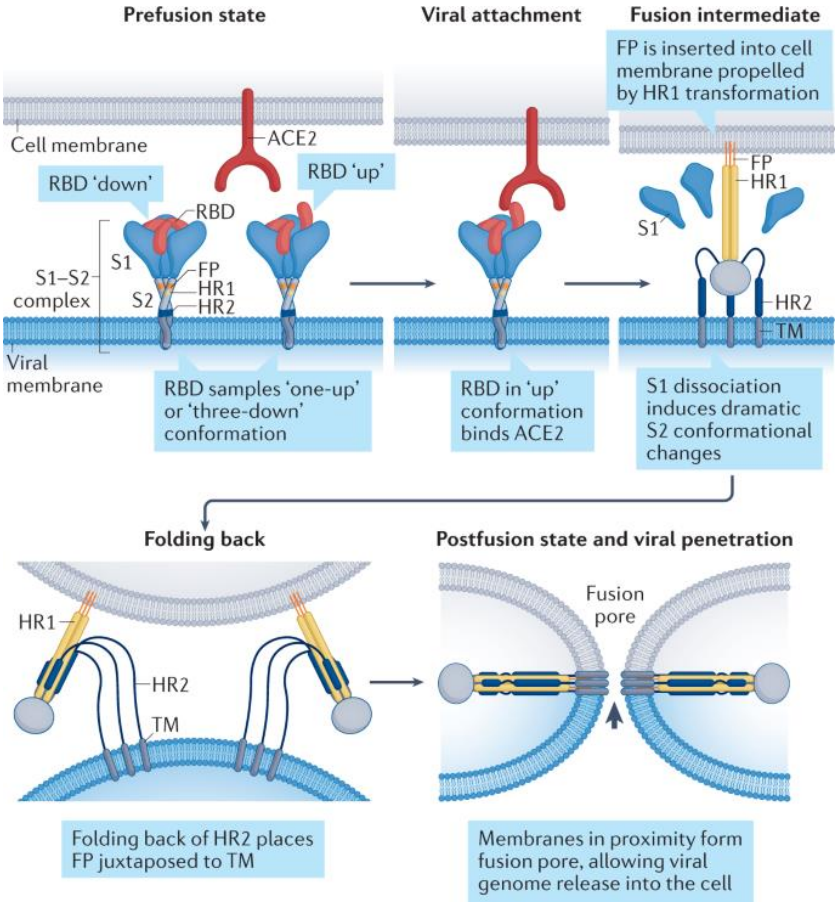


Fig I.9 membrane fusion induced by the SARS-CoV-2 S protein[73]

8.2. Viral Protein Translation Process

Upon the introduction of the coronavirus genome into the host cytoplasm, a sophisticated and finely orchestrated cascade of viral gene expression is initiated[83]. Fig I.10 illustrates all the infection cycle: The schematic illustrates the comprehensive life cycle of SARS-CoV-2. The initial step involves the binding of the virus's spike (S) glycoprotein to the ACE2 receptor on the host cell's surface, triggering endocytosis. Subsequently, the viral genomic RNA is released into the cytoplasm, initiating a cascade of events. The positive-sense genome undergoes transcription and translation, yielding nonstructural proteins (nsps), including key components like the RNA-dependent RNA polymerase and helicase. These proteins collaborate to form the RdRp complex. Within this complex, subgenomic transcription and RNA replication transpire, generating both negative-strand guide RNA (gRNA) and a collection of subgenomic RNAs crucial for viral replication and transcription. The newly formed subgenomic RNAs serve as templates for the synthesis of vital viral structural proteins, including spike (S), nucleocapsid (N), membrane (M), and envelope (E) proteins. The ensuing stages encompass the insertion of these structural proteins into the rough endoplasmic reticulum (ER) membrane. Subsequently, they are transported to the ER-Golgi intermediate compartment (ERGIC), where they unite with RNA encapsulated by the nucleocapsid (N) protein, culminating in the assembly of fully formed viral particles. The final phase of the cycle involves the release of these virions from the host cell through exocytosis, completing the virus's life cycle. More details of each step are explained below.

The gene expression process culminates in the synthesis of two pivotal polyproteins, pp1a and pp1ab, facilitated by the translation of Open Reading Frame 1a (ORF1a) and Open Reading Frame 1b (ORF1b) from the viral genomic RNA. It is noteworthy that the expression of pp1ab necessitates a programmed-1 ribosomal frameshift event positioned between ORF1a and ORF1b. The quantification of frameshift efficiency can be undertaken using ribosomal profiling techniques. In the case of SARS-CoV-2, the levels of pp1a expression surpass those of pp1ab by approximately 1.4–2.2 times[85].

Through a combination of co-translational and post-translational mechanisms, guided by viral proteases nsp3 and nsp5, sixteen non-structural proteins are liberated from pp1a (nsp1–11) and pp1ab (nsp1–10, nsp12–16). Significantly, fifteen of these proteins congregate to form the Replication and Transcription Complex (RTC), which plays an indispensable role in the virus's life cycle. This RTC harbors a suite of enzymes responsible for RNA processing, RNA modification, and RNA proofreading. Among these enzymes, the viral RNA-dependent RNA polymerase (RdRp), denoted as nsp12, collaborates harmoniously with its cofactors, nsp7 and nsp8, to execute RNA synthesis, concomitant with a 3′-terminal adenylyl-transferase activity. The bifunctional nature of nsp14 encompasses both a 3′ to 5′ exoribonuclease (ExoN) domain and a guanine-N7-methyltransferase (N7-MTase) domain. Previous studies underscore the critical role of nsp14's ExoN function in enhancing the fidelity of SARS-CoV replication over extended periods. Notably, recent biochemical analyses of MERS-CoV and SARS-CoV-2 mutants have unveiled an expanded role of ExoN beyond fidelity maintenance, involving a more direct contribution to RNA synthesis for these viruses[86].

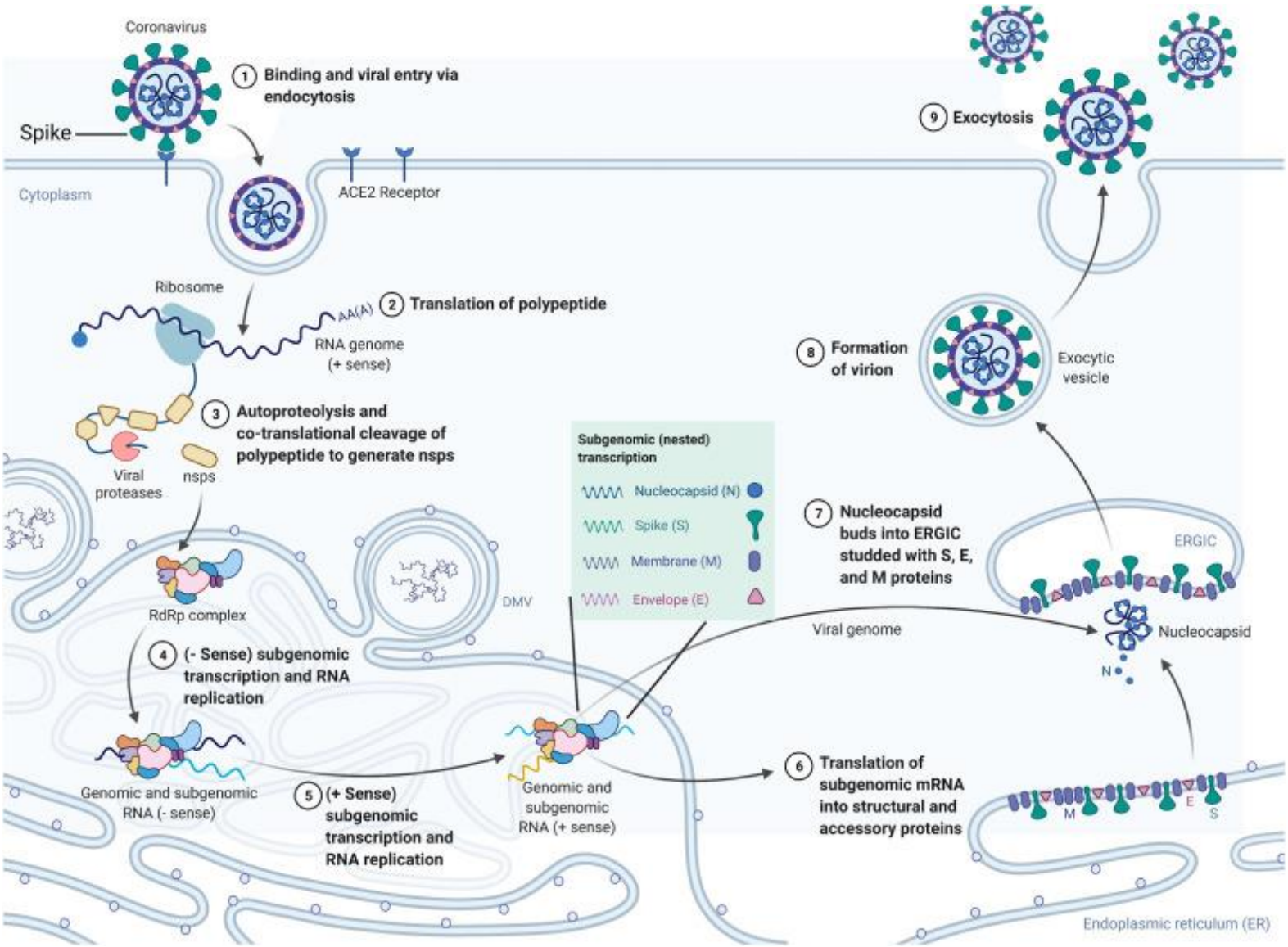


Fig I.10. Overview of the SARS-CoV-2 Life Cycle[84]

In addition to their enzymatic functions, non-structural proteins contribute essential ancillary roles to sustain the RTC's operations. These roles encompass modulating intracellular membranes, evading the host immune response, and providing cofactors that expedite the replication process. The final third of the SARS-CoV-2 genome is earmarked for genes encoding viral structural and accessory proteins. Among these, the spike protein (S), nucleocapsid protein (N), membrane protein (M), and envelope protein (E) play pivotal roles. For instance, the S1 subunit of the spike protein facilitates attachment to the ACE2 receptor, while the S2 subunit drives membrane fusion. Similarly, the N, M, and E proteins assume distinct responsibilities in various facets of viral functioning, maturation, and assembly, collectively orchestrating the virus's life cycle[87].

8.3. Variants of concern in Algeria

Phylogenetic analysis of SARS-CoV-2 in Algeria revealed valuable insights on the genomic diversity of the virus. The analyzed genomes formed fifteen distinct clades, including the Delta variant (AY.X, B.1.617.2), which is classified as a variant of concern (VOC) by the World Health Organization (WHO)[158]. As shown in Fig 11, the Nextclade analysis revealed that seventeen of the 36 SARS-CoV-2 genomes belonged to the Delta clade (21J), with the rest being part of clades 20A, 20B, and 20C[158].

Various mutations detected within Algerian sequences. Notable non-synonymous mutations were identified, including T85I in the nsp2 gene, P423L in the nsp12 gene, D614G in the S gene, and Q57H in the ORF3a gene[159]. Additionally, amino acid replacements characteristic of newly identified SARS-CoV-2 variants were found, including H69del, V70del, E484K, Y144del, and Q52R[159]. These mutations emerged after repatriation of Algerian nationals from abroad and were associated with heightened vigilance and lockdown measures[159].

Distinct mutations were also reported. For instance, the T1004I replacement in the nsp3 gene was unique to an Algerian sequence and had previously been reported in the early stages of the pandemic in the USA[159]. This connection suggested either travel history to the USA or contact with an individual linked to the USA-originated infection[159].

Furthermore, deleterious mutations that had functional implications were found. A deleterious amino acid replacement A130V in the RdRp gene and a non-synonymous replacement N874H in the NSP12 gene were detected. These mutations were observed in limited worldwide samples and could potentially impact viral replication and infectivity[159].

The diversity of viral genomes in Algeria can be categorized into five lineages (Fig I.11). Lineage A was considered the root of the pandemic, with sequences originating from China. Algerian sequences within this lineage exhibited associations with various clades, including B.1.1.7 (UK), B.1.351 (South Africa), and B.1.525 (Nigeria) related mutations[158,159].

8.4. BA.2.86 variant (Pirola): A new VOC?

Scientists are currently in a race to ascertain the potential global significance of a highly mutated coronavirus variant that has emerged on multiple continents. The variant, named BA.2.86 (or Pirola as some media outlets named it), was detected by various laboratories recently. Unlike other prevalent variants, BA.2.86 displays distinct characteristics and possesses numerous alterations in its spike protein, a crucial element in the immune response against the SARS-CoV-2 virus[160].

The emergence of BA.2.86 has evoked parallels to the appearance of the Omicron variant in late 2021, when an unusual lineage was identified in southern Africa and subsequently spread worldwide. The increased global immunity due to successive COVID-19 waves and booster vaccine campaigns suggests that BA.2.86 might not replicate the impact of Omicron. While expectations lean towards a less significant effect similar to the Omicron wave, it is still early in the assessment phase[161].

Efforts to comprehend BA.2.86's attributes and implications are explored. The variant, linked to six cases across four countries by August 21 2023, has garnered attention from the World Health Organization, which designated it as a 'variant under monitoring'. The lineage appears to have descended from the Omicron subvariant BA.2 but exhibits 34 distinct changes in its spike protein compared to BA.2. This significant mutation count likely originates from a chronic SARS-CoV-2 infection[161].

BA.2.86's distinctiveness has engendered close scientific observation. Unlike the anticipated pattern of SARS-CoV-2 evolution after Omicron's emergence, BA.2.86 is dramatically different from prevalent variants. This uniqueness, akin to the onset of Omicron and early pandemic variants like Alpha and Delta, has prompted vigilance from researchers[161].

Researchers are particularly intrigued by BA.2.86's impact on immune response. Numerous alterations in regions targeted by neutralizing antibodies suggest the variant's potential to elude immunity from past infections and vaccine boosters. Additionally, the variant's wide geographical distribution has captured attention, with unlinked cases indicating possible widespread transmission[161].

Current investigations focus on gauging BA.2.86's prevalence through patient samples and wastewater analysis. Isolation efforts and studies involving safe SARS-CoV-2 models aim to understand the variant's ability to evade neutralizing antibodies. Researchers are particularly keen on assessing the variant's interaction with antibodies triggered by the XBB.1.5 variant, the basis for the latest COVID-19 booster vaccines[161].

While vigilance is warranted, widespread concern about BA.2.86 is deemed unnecessary. Researchers emphasize that even if the variant becomes prevalent and adept at evading neutralizing antibodies, other forms of immunity should prevent severe illness in most individuals. The emergence of BA.2.86 underscores SARS-CoV-2's unpredictability and ongoing exploration into its characteristics and implications for public health[161].

8.5. What the future may hold

Looking ahead, the evolution of SARS-CoV-2 brings forth potential scenarios. Variants like Alpha and Delta were linked to higher transmissibility and some immune evasion. However,

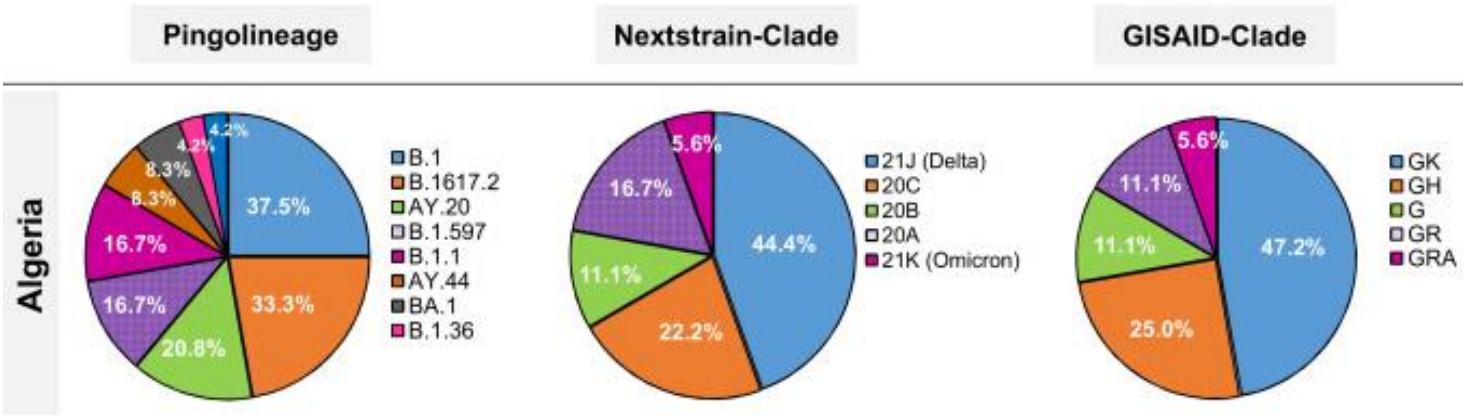


Fig I.11 Distribution of various SARS-CoV-2 clades and lineages from Algeria[159]

Omicron's rise was fueled by immune evasion, supplanting Delta. This suggests scenarios of co-circulating lineages or dominance of one, affecting public health[128,145]. Identifying key mutations early is possible, but their impact often hinges on others. Effective early warnings need comprehensive variant data and epidemiological insights. Wastewater surveillance aids in detecting hidden transmission. A positive scenario envisions drift within Omicron, combining immunity from prior infection and vaccination for protection. Yet, emergence of distinct variants from prolonged viral persistence in immunocompromised individuals is possible. These scenarios could impact virulence and public health outcomes[128,145]. SARS-CoV-2's future relies on emerging variants, vaccine development, and virulence evolution. Without eradication, the virus might become endemic with seasonal shifts. It's vital to remember endemicity doesn't guarantee safety; the pandemic-to-endemic shift is unpredictable. Adapting public health strategies according to evolving epidemiology is crucial. Despite growing understanding, the virus's evolution remains uncertain[128,145].

9. Vaccines

Following the release of the SARS-CoV-2 genome sequence on January 11th, 2020, an unparalleled effort in terms of both rapidity and scale was initiated to create a vaccine targeting COVID-19. According to WHO, as of 6:30pm CEST, 16 August 2023, a total of 13 499 865 692 vaccine doses have been administered, among which 5 149 830 832 persons vaccinated with a complete primary series. Fig I.12 illustrates the distribution of the number of vaccines per 100 people. Most countries in America, Europe, Asia, and Australia have vaccination coverage greater than 100, indicating that most people have administered a second dose. On the other hand, African countries have less vaccination rate. Madagascar, among others, has a very low vaccination rate of less than 20 doses administered per 100 population. Countries like Egypt, Tunisia, and Morocco have much higher vaccination coverage of >100. In Algeria, a total of 15 267 442 vaccine doses have been administered as of 4 September 2022, with a vaccination coverage ranging from 20 to 39 per 100 population. This variability in the vaccination coverage Within Africa can be due to different financial ability to buy vaccines and different vaccination policy in the absence of a unified approach in contrast to Europe (European Union).

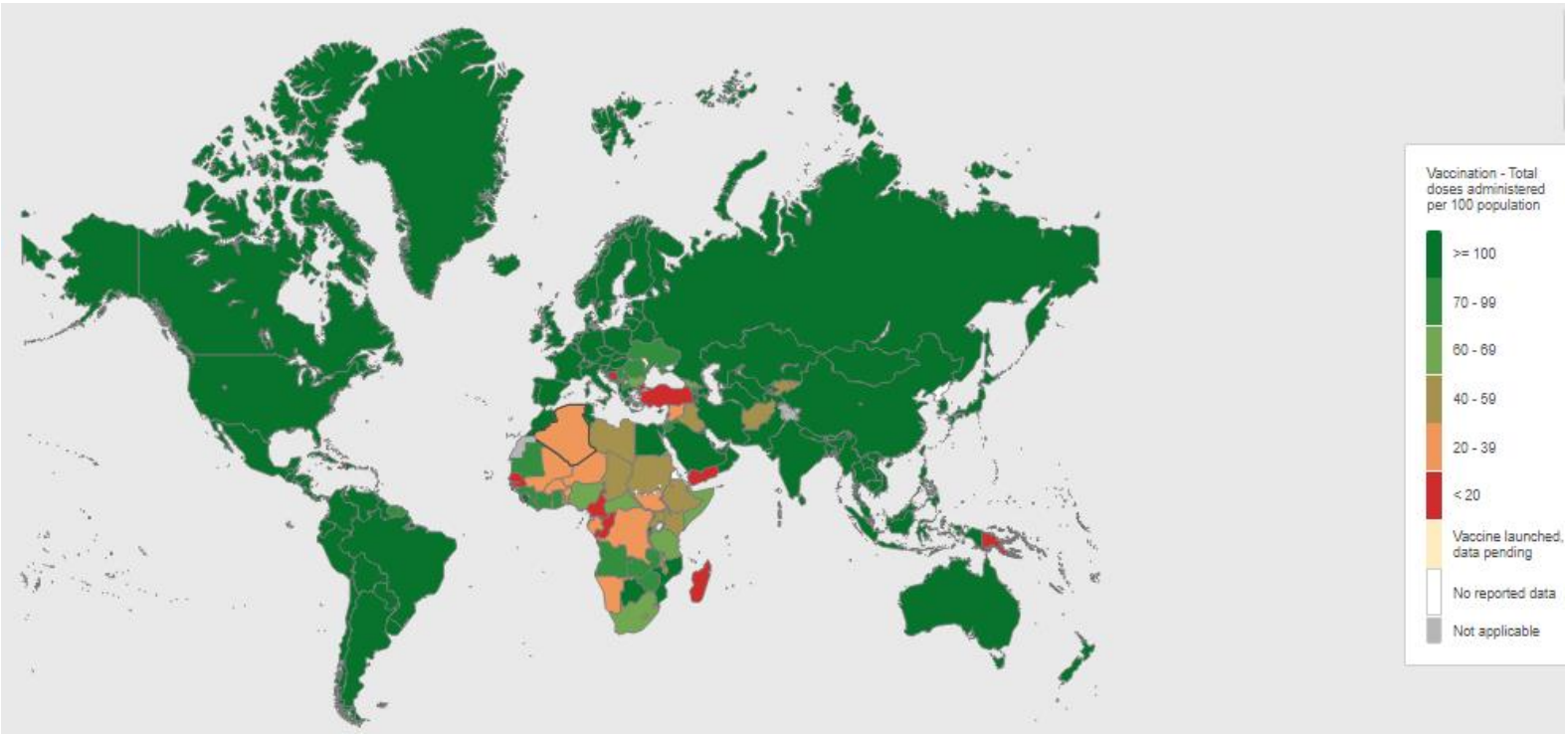


Fig I.12 Vaccination coverage in the world (WHO)

9.1. Approved vaccines

11.1.1. Nucleic acid vaccines: The mRNA-1273 vaccine candidate, a product of collaboration between Moderna and the US National Institute of Allergy and Infectious Diseases (NIAID), employs a small segment of genetic material called messenger RNA (mRNA) to direct cells to manufacture a protein located on the surface of the SARS-CoV-2 virus. This protein subsequently triggers an immune recognition, leading to the production of an immune response that offers potential protection against the virus[162]. Encouraging results had emerged from initial clinical trials, showcasing elevated levels of neutralizing antibodies and robust T cell reactions. Phase 3 trial involving 30,000 participants, this vaccine is administered in two doses spaced 28 days apart. Notably, safety and tolerability have been demonstrated[162].

11.1.2. Non-replicating viral vectored vaccines: AZD1222, originating from the University of Oxford and AstraZeneca, is a vaccine contender that employs a weakened form of an adenovirus (a common cold virus) to ferry genetic material from the SARS-CoV-2 virus into cells. This genetic information prompts cells to generate a protein found on the virus's surface, thereby prompting an immune response. Clinical trials have showcased promising outcomes, highlighting substantial levels of neutralizing antibodies and robust T cell responses. The Phase 3 trial encompassed 50,000 participants. It involves a two-dose regimen administered four weeks apart and has demonstrated safety and tolerability[162].

11.1.3. Inactivated virus vaccines: Developed by Sinovac Biotech, the CoronaVac vaccine candidate leverages inactivated SARS-CoV-2 virus to stimulate an immune reaction. Clinical trials have demonstrated promising results, with significant levels of neutralizing

antibodies. The Phase 3 trial involved 9,000 participants in Brazil. Administered in two doses, separated by a 14-day interval, the vaccine has displayed safety and tolerability[162].

11.1.4. Recombinant subunit vaccines: NVX-CoV2373, a vaccine contender from Novavax, utilizes a protein present on the surface of the SARS-CoV-2 virus to induce an immune response. Employing recombinant DNA technology, this involves introducing genetic material from the virus into cells for protein production. Encouraging outcomes from preliminary clinical trials include robust levels of neutralizing antibodies and noteworthy T cell responses. The vaccine's administration involves a multi-dose approach. It has demonstrated a positive safety and tolerability profile in early stages of assessment[162].

9.2. Vaccines Comparison: Pfizer, Moderna, and Johnson & Johnson

Pfizer and Moderna have developed vaccines utilizing messenger RNA (mRNA) technology that encodes the spike protein of the SARS-CoV-2 virus. On the other hand, Johnson & Johnson has formulated a vaccine based on an adenovirus vector platform[163]. Distinctively, the dosing regimen varies between these vaccines, as Pfizer and Moderna vaccines necessitate a two-dose administration, whereas the Johnson & Johnson vaccine is administered as a single dose. Noteworthy is the disparity in efficacy rates, with Pfizer and Moderna exhibiting an efficacy of 95.0% against moderate and symptomatic SARS-CoV-2 infection, while the Johnson & Johnson vaccine showcases an efficacy of 66.0%. Age-wise, Pfizer and Moderna vaccines are sanctioned for individuals aged 12 and above, whereas the Johnson & Johnson vaccine's authorization extends to those aged 18 and older. Storage requisites further differentiate these vaccines, with Pfizer and Moderna vaccines demanding ultra-cold storage, in contrast to the Johnson & Johnson vaccine which maintains stability under standard refrigeration temperatures[163].

Potential vaccine side effects were examined[163], predominantly characterized by mild to moderate manifestations that tend to resolve autonomously within a few days. The prevailing side effects for Pfizer and Moderna vaccines include localized pain and swelling at the injection site, fever, fatigue, headache, chills, and muscular as well as joint discomfort. Analogously, the Johnson & Johnson vaccine leads to analogous side effects such as injection site pain and swelling, headache, fatigue, muscle ache, and nausea. Severe adverse effects remain exceedingly rare occurrences[163].

As for the efficacy of these vaccines against emerging coronavirus variants, ongoing research continues to illuminate this facet. Notably, efficacy has demonstrated some decline against the Delta variant, which has been earmarked as a significant variant of concern by the World Health Organization. In a specific context, in Qatar, the vaccine's efficacy against B.1.617.2 infection was established at 64.2% fourteen days following the initial dose, and 53.5% fourteen days after the second dose, demonstrating its efficacy in averting hospitalization and mortality[164]. While its capacity to prevent infection was comparatively reduced, it remained effective in mitigating the more severe outcomes of hospitalization and fatality. Similarly, within the United Kingdom, efficacy was determined at 33.5% after the first dose and 59.8% after the second dose against the B.1.617.2 variant in preventing symptomatic infection[163].

10.Potential drugs

Approaches aimed at combating SARS-CoV-2 encompass a range of strategies such as the development of vaccines, repurposing existing drugs, designing novel antiviral agents, and therapeutic antibodies[84]. Consequently, certain pharmaceutical compounds have been meticulously designed to impede the formation of the viral capsid by engaging van der Waals forces, establishing hydrogen bonds, and orchestrating electrostatic interactions directly with the capsid structure. Conversely, other compounds exert their influence by modifying the intermolecular forces that uphold the integrity of the capsid, thereby disrupting its capacity to adhere to and infiltrate host cells[84]. Moreover, an array of small molecule inhibitors is directed towards several non-structural proteins, including the S-protein, M-protein, and RdRp, with a focus on binding to their respective enzyme active sites. Notably, some of these pharmaceutical agents are already accessible[84]. To illustrate, a multitude of in silico screening endeavors have revealed parallels between SARS-CoV-2 and other non-coronaviruses, thereby facilitating the repurposing of a variety of existing antiviral medications and drugs that are presently available in the market[84].

10.1. Novel drugs in phase III

1.1.1. PF-07321332(Paxlovid)

Paxlovid, a product of Pfizer, is an antiviral drug that combines nirmatrelvir, an inhibitor of the SARS-CoV-2 3CLpro responsible for viral growth, and ritonavir, a booster that enhances the efficacy of nirmatrelvir[165]. Initial tests of Paxlovid on SARS-CoV-1 showed promise, prompting evaluation against SARS-CoV-2. Laboratory experiments revealed encouraging results, with an 88% reduction in hospitalization and mortality among unvaccinated individuals. While a subsequent trial on vaccinated individuals with risk factors did not yield significant recovery acceleration, Paxlovid gained FDA approval for older SARS-CoV-2 patients based on its efficacy in reducing hospitalization risk[165].

1.1.2. S-217622

S-217622 is the first nonpeptidic, noncovalent, 3CLpro oral inhibitor candidate for treating COVID-19[166]. It was developed by by Shionogi [Osaka, Japan][167]. The prevailing inhibitors were based on peptide substrate mimics with covalent warheads that interacted covalently with Cys145 in the active site of 3CLpro. It was anticipated that these peptidic and reactive structural aspects could lead to challenges in the drug's pharmacokinetic properties, including oral bioavailability issues arising from poor cell permeability, metabolic instability, and susceptibility to degradation in the bloodstream. In response, a comprehensive search for nonpeptidic 3CLpro inhibitors was initiated using a structure-based drug design SBDD approach, aiming to address the ongoing SARS-CoV-2 pandemic[166].

S-217622 exhibited encouraging preclinical attributes as a once-daily oral therapy for COVID-19. It demonstrated potent antiviral activity against notable variants of concern, a prolonged elimination half-life in vivo, particularly in monkeys and dogs, excellent oral bioavailability, and marked efficacy in a mouse model infected with SARS-CoV-2. These promising characteristics prompted the advancement of S-217622 into clinical trials, which are currently ongoing[166].

1.1.3. AT-9010

AT-9010, an RdRp inhibitor developed by y Roche [Basel, Switzerland] and Atea [Boston, MA, USA][167], exhibits proficient integration into viral RNA, triggering immediate cessation of the RNA chain and demonstrating resilience against removal[168]. While the specific effects of AT-9010 on individual processes remain unclear, it effectively obstructs NMPylation activities (the enzymatic process of transferring a nucleotide monophosphate (NMP) group to a target molecule, typically a protein or a nucleic acid) involving nsp8 and nsp9, potentially impacting subsequent procedures like RNA capping, DNA/RNA ligation, and priming of RNA synthesis[168]. Notably, AT-9010's concurrent engagement with RdRp and NiRAN activities provides significant impetus for COVID-19 treatment, lessening the likelihood of simultaneous resistance mutations and aiding in managing the emergence of SARS-CoV-2 variants. This attribute gains special significance considering the genetic stability of SARS-CoV-2 throughout the ongoing pandemic[168].

10.2. Drug repurposing

Drug repurposing, a promising avenue in therapeutic agent development, explores novel applications for existing drugs[169,170]. This approach involves investigating established antiviral medications, originally intended for malaria, MERS, and SARS, for potential COVID-19 treatment. Some of these agents are currently undergoing clinical trials to assess their effectiveness against COVID-19 infection[171].

1.1.4. RNA-Dependent RNA Polymerase Inhibitors

Among the captivating drug targets for SARS-CoV-2 are the RNA-dependent RNA polymerase (RdRP) and a viral enzyme crucial for viral RNA replication within host cells. RdRP, being a viral enzyme distinct from host cell homologs, opens possibilities for developing selective inhibitors with heightened potency and minimal off-target effects. Notably, two nucleoside analogs, remdesivir and favipiravir, have been used for treating hospitalized COVID-19 patients[172].

Recent research identifies remdesivir as a potential COVID-19 treatment[173]. Functioning by inhibiting RNA polymerase and integrating into nascent viral RNA, remdesivir yields two significant outcomes: termination of the viral RNA chain and inhibition of viral genome replication. While remdesivir was found ineffective against Ebola, its human safety profile facilitated swift clinical trials for COVID-19 treatment and prevention[174]. Studies involving MERS-CoV infected mice demonstrated remdesivir's reduction of lung viral load, improved lung function, and reduced lung tissue damage[175]. Similar favorable effects were witnessed in both in vivo and in vitro tests against SARS-CoV and MERS-CoV[176].

Favipiravir, requiring activation through phosphoribosylation, emerged as a therapeutic option, gaining approval for managing novel influenza in China. It was introduced in Wuhan for treating SARS-CoV-2 infections and later approved for emergency use in various countries[177]. Studies indicated that patients administered favipiravir exhibited reduced pneumonia symptoms[178].

1.1.5. Ebselen

Ebselen, previously recognized for its antiviral potency against diverse viruses, including HIV-1, HCV, influenza A, and Zika virus, has demonstrated promise in mitigating liver injury triggered by chemical and microbial factors. Given the correlation between severe COVID-19 cases and liver injury, ebselen's potential benefits are notable[177]. Furthermore, ebselen's efficacy in managing focal ischemic injury and reducing IL-6 suggests potential benefits in safeguarding SARS-CoV-2 infected patients from venous thrombosis and vascular injury[179]. However, conclusive evidence necessitates randomized clinical trials to substantiate its incorporation into treatment regimens[180].

1.1.6. Arbidol

Arbidol's broad-spectrum inhibition against parainfluenza, influenza A and B viruses, and hepatitis C virus has led to its consideration as a preventive agent among health professionals[177]. Notably, its synergistic effect with interferons enhances its therapeutic potential against coronaviruses[181].

1.1.7. Interferons

Interferons, recognized for their antiviral properties, have garnered interest for managing COVID-19. Interferon-α, with its broad-spectrum antiviral activity, is particularly notable. While it has shown effectiveness against SARS-CoV replication, its partial efficacy against coronavirus highlights the complexities of its antiviral mechanisms[182,183].

1.1.8. Protease Inhibitor

Protease inhibitors, which are vital for viral replication within host cells, hold potential for treating severe acute respiratory syndrome virus 2. Lopinavir/ritonavir, known for managing HIV, have been explored for COVID-19 treatment due to their effectiveness against MERS[177]. Studies reveal their capability to inhibit COVID-19's proprotease and their effectiveness against MERS and SARS viruses[184,185]. Though exhibiting some advantage in COVID-19 patient outcomes, the effectiveness of lopinavir/ritonavir combination remains modest[177].

1.1.9. Ivermectin

Ivermectin, with its broad-spectrum antiviral activity against various viruses, exhibits potential by targeting viral nuclear transport processes, particularly significant for RNA viruses[186].

1.1.10. Melatonin

Melatonin, recognized for its immune-modulatory role, plays a pivotal role in attenuating cytokine storms associated with severe COVID-19 cases, thus reducing mortality and morbidity[187].

1.1.11. Teicoplanin

Teicoplanin, known for inhibiting viral replication through the cleavage of viral spike protein, holds potential as a therapeutic agent for COVID-19 treatment[188].

1.1.12. 5-Alpha Reductase Inhibitors

Finasteride and dutasteride, 5-alpha reductase inhibitors, offer promise in mitigating severe COVID-19 by altering hormone-related factors[177].

1.1.13. Spironolactone

Spironolactone, an antihypertensive with potential antiandrogenic effects, presents opportunities for safeguarding against COVID-19 by addressing hormone-related factors[177].

1.1.14. Sildenafil/Tadalafil

PDE5 inhibitors like sildenafil and tadalafil, under investigation for COVID-19, show promise in preventing fibrosis and improving respiratory parameters[189].

1.1.15. Statins

Statins, recognized for anti-inflammatory effects, are being studied for their potential in improving outcomes in severe COVID-19 cases[177].

1.1.16. N-Acetylcysteine

N-Acetylcysteine, known for its antioxidant and immune-modulatory effects, has potential in inhibiting viral replication through negative regulation of Nuclear Factor-kappa B NF-κB[190].

1.1.17. Bromhexine Hydrochloride

Bromhexine hydrochloride, an over-the-counter expectorant, presents as a potential COVID-19 treatment candidate based on its efficacy against TMPRSS2 protease[191].

1.1.18. Chloroquine Phosphate or Hydroxychloroquine

Chloroquine and hydroxychloroquine, with a history of antiviral effects, are being evaluated for COVID-19 treatment. Their mechanism involves hindering viral binding and replication[192,193].

1.1.19. Corticosteroids

Corticosteroids, despite concerns about immunosuppression, have been effective in reducing mortality rates in severe COVID-19 cases[177].

10.3. Traditional medicine

The utilization of traditional medicinal practices to manage infectious diseases has witnessed a surge, driven by several factors, including the involvement of various concerned entities in the development and investigation of herbal-based remedies[194,195]. Recent investigations have delved into the therapeutic potential of herbal medicines for combating COVID-19[196]. Historical precedent exists for the use of plant-derived herbal remedies in addressing various outbreaks, such as H1N1 influenza and SARS[197,198]. Presently, nations such as South Korea and China have endorsed guidelines pertaining to the use of traditional medicine in the context of COVID-19 management[199]. Notably, in China, a substantial eighty-five percent of COVID-19 patients have received traditional medicinal interventions[196]. A recent study has unveiled that traditional remedies have the potential to competitively target the angiotensin-converting enzyme 2 receptor, akin to SARS-CoV-2 and SARS-CoV, thus

opening avenues for leveraging traditional medicine in managing SARS-CoV-2 infections[200].

Among the array of plant-based traditional medicines employed for managing COVID-19 in China, notable instances include *Radix platycodonis*, *Agastache rugosa*, *Saposhnikoviae divaricata*, *Lonicerae japonicae flos*, *Astragalus membranaceus*, *Rhizoma Atractylodis Macrocephalae*, *Glycyrrhiza uralensis*, *Atractylodis Rhizoma*, *Fructus forsythia*, and *Cyrtomium fortunei J. Sm*[201].

11. Conclusion

Research and development of potential drugs and treatments for COVID-19 is still ongoing worldwide. However, traditional drug design and discovery techniques are time-consuming, extremely expensive, and usually fail in clinical trials. This is why Computer-Aided Drug Design CADD has emerged in recent year as a preliminary step to the wet lab experiments, taking advantage of advances in Artificial Intelligence AI and in computational hardware. The next chapter discusses CADD most famous techniques and their applications in identifying and discovering drugs for SARS-CoV-2.

CHAPTER II
Computer-Aided Drug Design

1. Introduction

In the ever-evolving landscape of pharmaceutical research and drug development, Computer-Aided Drug Design CADD has emerged as a powerful and indispensable tool[202]. This chapter delves into three pivotal aspects of CADD: Structure-Based Drug Design SBDD, Ligand-Based Drug Design LBDD, and Virtual Screening VS. These methodologies, underpinned by computational algorithms and simulations, have revolutionized the way we identify and optimize potential drug candidates[202]. By harnessing the synergy between cutting-edge technology and molecular insights, CADD has significantly accelerated the drug discovery process, making it more efficient, cost-effective, and tailored to the specific needs of modern medicine[203–208]. This chapter explores the principles, techniques, and applications of SBDD, LBDD, and VS, as these computational approaches continue to shape the future of pharmaceutical innovation.

2. Computer-aided drug design

The process of drug discovery is an extensive endeavor, spanning approximately 10-15 years[209], and incurring costs of up to 2.558 billion USD to bring a drug to market[210]. This intricate process involves several stages, commencing with the identification of a suitable drug target, followed by the validation of said target, progressing through hit-to-lead discovery, lead molecule optimization, and culminating in preclinical and clinical investigations[202]. Despite substantial investments in time and resources for new drug discovery, the success rate in clinical trials stands at a mere 13%, coupled with a notable rate of drug attrition[202]. In a significant proportion of instances (40-60%), later-stage drug failures can be attributed to suboptimal pharmacokinetic attributes pertaining to absorption, distribution, metabolism, excretion, and toxicity ADMET[202].

Computer-Aided Drug Design (CADD) refers to the interdisciplinary field that integrates computational techniques and methodologies into the process of drug discovery and development. It involves the utilization of computational tools to predict and model interactions between drug molecules and biological targets, thereby facilitating the identification of potential drug candidates[43]. The integration of computer-aided drug discovery (CADD) techniques has proven pivotal in expediting the drug development process and mitigating costs and final-stage failures[208,211]. Notably, rational drug design, an integral facet of CADD, furnishes valuable insights into the binding affinity and molecular interactions between target proteins and ligands. The provision of supercomputing capabilities, parallel processing, advanced algorithms, and tools has facilitated lead identification in pharmaceutical research[212]. Additionally, recent strides in artificial intelligence (AI) and machine learning methodologies have significantly enhanced the analysis, comprehension, and interpretation of pharmaceutical-related big data within the realm of drug discovery[213].

Various methodologies, such as pharmacophore modeling, quantitative structure-activity relationship (QSAR), molecular docking, quantum mechanics, and statistical learning, are harnessed for identifying new inhibitors from chemical databases[202]. CADD encompasses two predominant approaches: structure-based and ligand-based drug design. Both

methodologies find widespread application in identifying suitable lead molecules. The former relies on the three-dimensional structure of the target receptor and its active sites to elucidate molecular interactions with ligands[202]. Conversely, ligand-based drug design hinges on knowledge of ligand interactions with a given target receptor. The realm of computer-aided drug design boasts numerous success stories and continues to be a pivotal force in the drug discovery domain[202].

Nonstructural and structural proteins of SARS-CoV-2 present promising avenues for designing antiviral agents against COVID-19[202]. Given the elevated mortality rate COVID-19, the expeditious discovery of novel drugs is imperative[43]. Here, computer-aided drug design emerges as a critical tool for identifying novel therapeutics. The prospect of harnessing natural lead molecules obtained through virtual screening and pharmacokinetic prediction holds promise for developing effective lead molecules against COVID-19[214]. In the race to identify potential treatments for SARS-CoV-2 infection, repurposing broad-spectrum antiviral drugs proves a viable strategy due to the availability of pharmacokinetic and pharmacodynamic data for these agents[214].

1. Structure-based drug design

Structure-Based Drug Design (SBDD) hinges on the three-dimensional structure of therapeutic target proteins and exploration of their binding site cavities[202]. This precise and rapid approach aids in identifying and optimizing lead molecules, contributing to a profound understanding of diseases at the molecular level[202]. SBDD encompasses techniques such as structure-based virtual screening (SBVS), molecular docking, and molecular dynamics (MD) simulations, serving diverse purposes like evaluating binding energetics, protein-ligand interactions, and receptor conformational changes upon ligand binding[202].The workflow of SBDD is illustrated in Fig II.1. It has significantly aided pharmaceutical industries and medicinal chemists in discovering multiple drugs, such as amprenavir, raltitrexed, norfloxacin, dorzolamide, isoniazid, and flurbiprofen[202]. The SBDD process encompasses target structure preparation, ligand binding site identification, compound library curation, molecular docking with scoring functions, MD simulation, and binding free energy calculation[202].

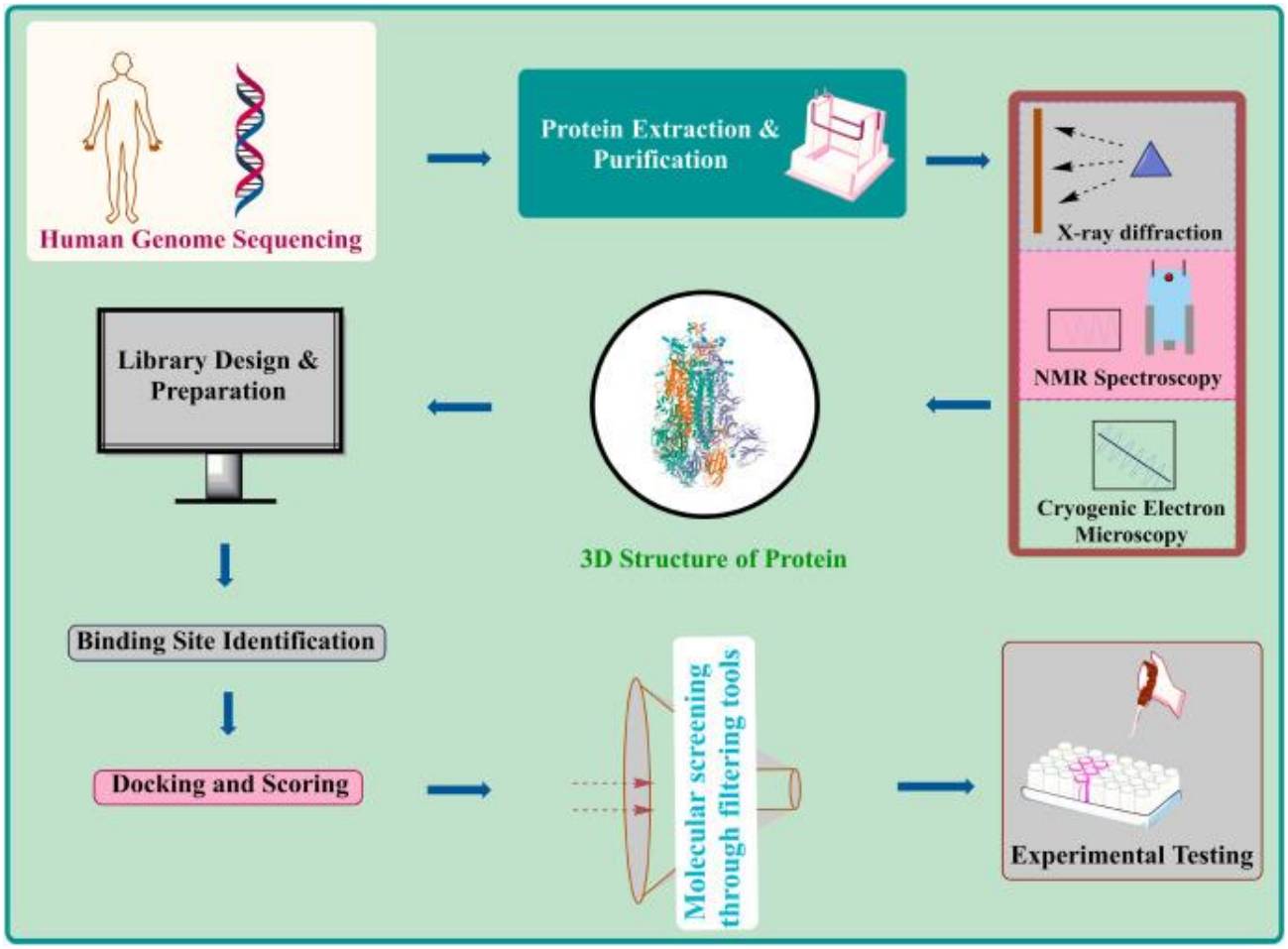


Fig II.1 SBDD workflow[215]

1.1. Molecular docking

Molecular docking is a computational approach employed for the examination of interactions occurring at the molecular level between a target receptor and a ligand. It facilitates the ranking of ligands by evaluating their binding affinity to the receptor, employing diverse scoring functions[216]. Achieving favorable binding poses of ligands within the active site of the target involves considering two critical factors: (a) exploring a broad conformational space that accounts for various binding orientations and (b) making explicit predictions about the binding affinity of ligands associated with each binding pose[217]. Table II.1 provides a list of commonly utilized molecular docking software programs. Molecular docking can be categorized into two main types: flexible-ligand search docking and flexible-protein docking. In the flexible-ligand search docking method, ligand flexibility is typically addressed through three algorithms: systematic, stochastic, and simulation methods[218]. On the other hand, flexible-protein docking primarily relies on Monte Carlo (MC) and molecular dynamics (MD) approaches[202]. Extensive explanation and details can be found in specialized literature[216–224].

Table II.1 Docking softwares

Tool	Key Features	Reference
AutoDock	Conformational searching methods: Lamarckian genetic algorithm, simulated annealing, traditional genetic algorithm. Prediction of binding free energies based on a semiempirical free energy force field.	[225]
AutoDock Vina	Utilizes gradient optimization for improved speed and accuracy in predicting binding modes compared to AutoDock.	[226]
GOLD	Automated ligand docking with ligand conformational flexibility and partial protein flexibility using a genetic algorithm.	[227]
CDOCKER	Automated MD docking program with CHARMm19 force fields, offering full ligand flexibility and reduced computation time.	[228]
FlexX	Automated docking tool for flexible ligands, relies on base fragment selection and placement for scoring.	[229]
Surflex	Uses Hammerhead's empirical scoring function and molecular similarity to produce ligand fragment poses.	[230]
GLIDE	Grid-based ligand docking with comprehensive search of ligand space, scoring based on the ChemScore function.	[231]
DOCK6	Docking program utilizing anchor-and-grow search algorithm for conformational sampling of small molecules.	[232]
SwissDock	Web server for docking small molecules to target proteins using the EADock DSS engine.	[233]
MOE	Molecular Operating Environment: Software package that provides a wide range of tools and functionalities for tasks such as molecular modeling, protein-ligand docking, structure-based design, cheminformatics, and more.	[234]
Molegro Virtual Docker	Software package with a user-friendly interface, several docking protocols, and post docking analyses	[235]
Schrodinger Suite	Software package that provides diverse tools, high accuracy, user-friendliness, interdisciplinary applicability, and a large user community.	[236]

1.2. Molecular docking procedure

1.1.20. Macromolecule preparation

The proliferation of structural elucidation methods like X-ray and NMR has led to a substantial increase in the number of protein structures available in databases like the Protein Data Bank (PDB) over recent decades. However, due to the inherent limitations of experimental techniques, there are still many target protein structures that remain unsolved[237].

To bridge this gap, computational techniques such as comparative homology modeling, threading, and ab initio modeling have emerged as successful means of deciphering protein structures from their amino acid sequences. One widely utilized computational method is homology modeling, which accurately determines a protein's three-dimensional structure by leveraging an appropriate template structure identified from existing data[238]. This process involves several steps, including template selection, sequence alignment, target model construction, model refinement, and validation[239].

Protein threading is another valuable approach for predicting protein structures, especially when the target protein exhibits limited sequence similarity with proteins in the PDB (less than 25% sequence identity) but shares structural similarities with some existing proteins. Unlike homology modeling, protein threading considers structural information, such as secondary structure, solvent accessibility, and pairwise interactions encoded in the template, enhancing prediction accuracy[240].

In cases where there are no template structures available in existing biological databases, ab initio modeling comes into play. This computational technique is preferred for predicting the structures of proteins with no existing templates. It involves addressing a global optimization problem to determine dihedral angle values that contribute to the stability of the protein structure, aiming to achieve the global or near-global minimum potential energy[241].

Subsequently, the structure undergoes rigorous cleaning, involving the removal of extraneous elements such as water molecules, cofactors, and ions that are unrelated to the binding site. Addressing potential structural deficiencies is the next priority, as missing atoms, residues, or side chains are repaired to create a complete and accurate representation. Adjusting protonation states of ionizable residues to match the intended pH conditions is crucial for realistic binding energy calculations[242].

Energy minimization steps follow to relax steric clashes and rectify unfavorable bond geometries, leading to a more stable protein conformation. Hydrogen atoms are meticulously placed, taking into account tautomeric and ionization states of amino acid residues. The binding site, where ligands are anticipated to interact, is carefully defined. A 3D grid is generated around this site to expedite the search for ligand binding poses, a feature employed by some docking programs. If any co-crystallized ligands are present, they may be removed unless re-docking studies are intended for validation.[243,244]. Despite the absence of

experimental data regarding the binding sites of numerous proteins, various software and webservers, including CASTp [36], DoGSite Scorer [37], NSiteMatch [38], DEPTH [39], MSPocket [40], MetaPocket [41], and Q-SiteFinder [42], enable the prediction of potential binding sites on target proteins. During the lead identification process, compounds with large sizes that do not readily fit into the binding site pocket are typically excluded[202].

Ultimately, this rigorous protein preparation process provides a sound foundation for conducting accurate and reliable molecular docking studies, making it an essential component of drug discovery and molecular research[242].

1.1.21. Small molecules preparation

Chemical compounds can be sourced from various chemical databases, including ZINC15 (comprising 230 million purchasable compounds)[245], MCULE (offering 122 million synthetically accessible compounds) (https://mcule.com/), PubChem (with 111 million pure and characterized chemical compounds)[246], ChemSpider (housing 25 million unique chemical compounds)[247], DrugBank (encompassing 14,528 drug molecules)[248], and ChEMBL (featuring over 1.6 million distinct compounds)[249].

In the context of molecular docking, the selection of compounds focuses on those with drug-like characteristics, filtered based on criteria such as Lipinski's rule and ADMET parameters. Additionally, risk factors like acute rat toxicity, carcinogenicity, serum glutamic oxaloacetic transaminase elevation, hepatotoxicity, and inhibition of 3A4 oxidation of midazolam are considered[237]. According to Lipinski's rule of five, a compound is deemed orally bioactive if its physicochemical properties fall within permissible limits defined by criteria such as molecular weight (≤500), logP (partition coefficient between n-octanol and water, ≤5), hydrogen bond donors (≤5), and hydrogen bond acceptors (≤10)[250,251].

Furthermore, ADMET properties of interest encompass human gastrointestinal absorption (HIA), blood-brain barrier (BBB) permeation, P-glycoprotein (P-gp) inhibition, cytochrome P450 (CYP) inhibition, and plasma protein binding. Alongside pharmacokinetic properties and safety, the feasibility of synthesizing these compounds (synthetic score) should also be taken into consideration[250].

1.3. Molecular Dynamic MD Simulation

In the late 1970s, the first Molecular Dynamics (MD) simulation of a protein was conducted[252]. This robust technique involves predicting the positions of every atom within a molecular system over time, based on Newton's laws governing interatomic interactions[253]. The forces between atoms are determined using a relevant force field, which in turn calculates the system's overall energy[254]. MD simulations offer several advantages. They provide a comprehensive record of the position and motion of each atom throughout the simulation, something challenging to achieve experimentally. Moreover, the simulation conditions are precisely controlled and adjustable[255]. MD simulations have found extensive use in structure-based drug discovery because they offer insights into intricate details like binding, unbinding, and receptor conformational changes at a high level

of detail, which is often inaccessible through experimental methods[202]. Additionally, MD simulations enable the exploration of receptor-ligand dynamics, including association, dissociation, and the quantification of thermodynamics, kinetics, and the free energy landscape[256]. Several MD simulation programs are available, including GROMACS[257], AMBER[258], CHARMM[259], NAMD[260], and Tinker[261]. The details of MD has been extensively explained in literature[252–255,258,262–265].

1.4. ADMET properties

ADMET, which stands for Absorption, Distribution, Metabolism, Excretion, and Toxicity, are crucial properties considered in the field of drug discovery and development[266]. These properties play a pivotal role in determining the efficacy and safety of potential drug candidates[266].

1.1.22. Absorption

This property refers to how a drug is taken up by the body after administration. It involves understanding how well a drug is absorbed into the bloodstream from its delivery route (e.g., oral ingestion, intravenous injection). Factors such as solubility, bioavailability, and permeability influence drug absorption[267].

1.1.23. Distribution

Once a drug is absorbed, it needs to reach its target site in the body. Distribution entails how the drug spreads throughout different tissues and organs. Factors such as blood flow, protein binding, and tissue permeability affect drug distribution[267].

1.1.24. Metabolism

The body's metabolic processes can alter the chemical structure of a drug. Metabolism primarily occurs in the liver and involves enzymatic reactions that can either activate or deactivate a drug. Understanding drug metabolism is crucial to determine its stability and potential for toxic byproducts[267].

1.1.25. Excretion

This property involves the removal of a drug and its metabolites from the body. Most drugs are excreted through the kidneys via urine, but other routes like feces and exhalation also play a role. Knowledge of excretion pathways aids in determining the drug's elimination half-life[267].

1.1.26. Toxicity

Assessing the potential adverse effects of a drug is essential for its safety profile. Toxicity studies investigate whether a drug or its metabolites can harm the body's cells or organs. This includes evaluating acute and chronic toxicity, genotoxicity, and carcinogenicity[267].

ADMET properties are assessed through various in vitro and in vivo experiments, computational modeling, and predictive tools. A thorough understanding of these properties is integral to identifying promising drug candidates and optimizing their efficacy while minimizing potential side effects, ultimately advancing the field of drug design and development. ADMET properties refer to a set of pharmacokinetic properties that are

important to consider in drug discovery and development. Effective and safe drugs need to have a finely tuned combination of pharmacodynamics (PD) and pharmacokinetics (PK), including high potency, affinity, and selectivity against the molecular target, along with adequate ADMET properties. In silico prediction of ADMET is an important component of pharmaceutical R&D[268]. Bad ADMET profiles are the primary causes of failure for the majority of drug candidates during clinical trials[269]. A substantial challenge in the realm of drug discovery and design is the holistic enhancement of ADMET properties[268]. Consequently, the utilization of in silico prediction methods for ADMET properties serves a dual purpose: not only does it enable the assessment of pharmacokinetics and pharmacodynamics attributes of small molecules, but it also aids in the identification of compounds with drug-like characteristics and potential as lead compounds[270]. Various web servers are available to predict ADMET, the most commonly used are ADMETLAB[271], Swiss ADME[250], and PKCSM[272] (Pharmacokinetics and Chemistry Score Matrix). ADMET properties and their interpretations have been extensively detailed in literature[273–281].

2. SBDD literature review

Das *et al*, (2021) .[282] used a blind molecular docking approach to identify potential inhibitors for the SARS-CoV-2 main protease. The screening focused on 33 diverse molecules, ranging from natural compounds to antivirals and anti-protozoals. Notably, all of these molecules exhibited binding capabilities within the active site of the main protease. Rutin, a natural compound, displayed the highest inhibitor efficiency, followed by ritonavir, emetine, hesperidin, lopinavir, and indinavir. These interactions mainly occurred near essential catalytic residues, such as HIS41 and CYS145, and involved other critical active site residues.

Deshpande *et al*, (2020) .[283] applied molecular docking analyses on key viral propagation-related proteins, including 3Clpro, Nsp10/16, Spike protein, SARS protein receptor binding domain, Nsp 9 viral single-strand binding protein, and viral helicase. Additionally, the human angiotensin-converting enzyme 2 (ACE2) receptor, which serves as the entry point for the virus into host cells, was also subjected to molecular docking experiments. The compounds used for docking include well-known drugs for SARS-CoV-2 such as Ritonavir, Lopinavir, Remdesivir, Chloroquine, Hydroxychloroquine (HCQ), as well as conventional antiviral medications like Oseltamivir and Ribavirin. Furthermore, the evaluation of small molecules with anti-inflammatory properties, such as Mycophenolic acid (MPA), Pemirolast, Isoniazid, and Eriodictyol was elaborated. The results highlight the therapeutic potential of Ritonavir, Lopinavir, and Remdesivir against SARS-CoV-2. Notably, Eriodictyol exhibited strong binding to almost all selected target proteins, indicating its potential significance in COVID-19 treatment. Molecular interactions observed between Ritonavir, Lopinavir, Remdesivir, and SARS-CoV-2 proteins further enhance their candidacy as potential drugs for COVID-19 treatment. Eriodictyol is emerging as a promising repurposed drug for potential use against COVID-19.

Kerkour *et al,* (2023) .[284] in their study, molecular docking analyses were carried out to assess the interactions between the DHPAMP ligand and SARS-CoV-2 Mpro and RdRpreceptors. The results revealed strong binding affinities, with DHPAMP displaying negative docking scores of -6.40 kcal/mol and -8.40 kcal/mol for Mpro and RdRp, respectively. These findings indicated robust binding interactions. Notably, DHPAMP established specific interactions, including van der Waals forces and hydrogen bonds, with key amino acids within the active sites of Mpro and RdRp. Comparative analysis of binding energies highlighted DHPAMP's superior affinity for these proteins, suggesting its potential as a promising inhibitor in the fight against COVID-19.

Mhatre *et al,* (2021) .[285] performed a comparative analysis of docking scores and interaction types, where EGCG and TF3 exhibited favorable docking scores (-8.3 kcal/mol and -8.4 kcal/mol, respectively, for the 3CLpro receptor). and demonstrated significant in-silico activity against potential COVID-19 targets.

Vivek-Ananth *et al* .[286] identified potential natural product inhibitors of SARS-CoV-2 Nsp13 targeting the ATP-binding site. Among the 368 potential phytochemical inhibitors, the top candidates for further study include Picrasidine M, (+)-Epiexcelsin, Isorhoeadine, Euphorbetin, and Picrasidine N. These top candidates, including Picrasidine M, (+)-Epiexcelsin, Isorhoeadine, Euphorbetin, and Picrasidine N, demonstrated stable binding interactions with the ATP-binding site of SARS-CoV-2 Nsp13.

Ghosh *et al,* (2021) .[287] selected eight polyphenols from green tea to identify potent Mpro inhibitors. The binding affinities and binding modes between these polyphenols, including the established Mpro inhibitor N3 (with a binding affinity of -7.0 kcal/mol), and Mpro were elucidated through molecular docking studies. All eight polyphenols showed strong interactions with one or both catalytic residues (His41 and Cys145) of Mpro. Molecular dynamics simulations (100 ns) on these three Mpro-polyphenol systems further revealed high stability, minimal conformational fluctuations, and consistent compactness. Estimation of the total number of intermolecular H-bonds and MM-GBSA analysis confirmed the stability of these three Mpro-polyphenol complexes. Pharmacokinetic analysis further suggested that these polyphenols exhibited favorable drug similarity characteristics.

Hall Jr *et al,* (2020) .[288] applied molecular docking on existing drugs and found that Zanamivir, Indinavir, Saquinavir, and Remdesivir are promising candidates for inhibiting the 3CLPRO main proteinase. Additionally, they uncovered that Flavin Adenine Dinucleotide (FAD), Adeflavin (a B2 deficiency medicine), and Coenzyme A, a coenzyme, could also hold promise for treating SARS-CoV-2 infections.

Li *et al,* (2021) .[289] employed molecular docking approach to identify potential inhibitors of the SARS-CoV-2 papain-like protease. The docking involved 21 compounds with antiviral, antifungal, and anticancer properties. Remarkably, Neobavaisoflavone exhibited the highest binding affinity for the SARS-CoV-2 papain-like protease. These compounds demonstrated the capability to bind in proximity to crucial catalytic residues of the papain-like protease, including Trp106, Asn109, Cys111, Met208, Lys232, Pro247, Tyr268, Gln269, His272, Asp286, and Thr301.

Rahman *et al*, (2021) .[290] evaluated the potential of rutin, a medicinal flavonoid, to inhibit SARS-CoV-2 by molecular docking. Rutin demonstrated notable binding interactions with multiple vital proteins of SARS-CoV-2, including Mpro, RdRp, PLpro, and S-proteins. Among these, Mpro exhibited the most robust binding affinity with the lowest binding energy (-8.9 kcal/mol). This interaction was stabilized through hydrogen bonds with bond lengths ranging from 1.18 Å to 3.17 Å, along with hydrophobic interactions. Evaluation of ADMET and bioactivity predicted optimal solubility, non-toxicity, and non-carcinogenic properties for rutin. The predicted inhibitory constant values for rutin with these essential SARS-CoV-2 proteins fell within the range of 5.66 μM to 6.54 μM.

EL Khatabi *et al,* (2022) .[291] applied molecular docking and ADMET on Levamisole derivatives. Among them, molecule N°6 exhibited particularly encouraging docking outcomes, displaying favorable binding interactions and superior docking energy when compared to chloroquine and mefloquine. The results of ADMET prediction and MD simulation provide further evidence of the potential of molecule N°6 as a promising candidate for further development as a novel inhibitor capable of combatting the newly emerged SARS-CoV-2.

Choudhary *et al,* (2020) .[292] applied molecular docking and molecular dynamics simulations on 9 compounds susceptible of inhibiting 3CLpro. 2 compounds showed the best binding energy of -71.63 kcal/mol and -81.92 kcal/mol, and were confirmed to form stable complexes by molecular dynamics simulation of 20 ns simulation period.

Wahedi *et al*, (2021) .[293] reported promising potential in repurposing stilbenoid analogs, particularly resveratrol, as anti-COVID-19 drug candidates. Molecular docking analysis revealed that all four tested compounds had good binding affinity to the SARS-CoV-2 spike protein and human ACE2 receptor complex, with docking scores exceeding -7 kcal/mol. Furthermore, extensive molecular dynamics simulations conducted over fifty nanoseconds in aqueous solution demonstrated that resveratrol exhibited a highly stable bound conformation when interacting with the viral protein-ACE2 receptor complex. Additionally, the net free energy of binding calculated confirmed the stability of the resveratrol-protein complex.

Al-Karmalawy *et al,* (2021) .[294] used molecular docking and dynamics simulations and identified 2 potential molecules. Alacepril and lisinopril demonstrated a noteworthy interaction with human angiotensin-converting enzyme 2 (hACE2), which serves as the host's entry point for the SARS-CoV-2 spike protein. This interaction was substantiated by their favorable rmsd_refine values and superior binding affinity. Notably, both compounds formed robust hydrogen bonds with Asn90, a residue believed to be pivotal for receptor activity. Additionally, these ligands exhibited significant supplementary interactions with other receptor-binding residues. MD simulations demonstrated good stability with the target.

Kushwaha *et al,* (2021) .[295] applied molecular docking, MD simulation, and ADMET properties on 130 phytochemicals against three crucial SARS-CoV-2 targets: Spike (S)-protein, RNA-dependent RNA polymerase (RdRp), and Main protease (Mpro). Notably, Isoquercetin exhibited remarkable binding potential with the active sites of Spike, RdRP, and Mpro, yielding docking scores of -8.22, -6.86, and -9.73 kcal/mole, respectively.

Additionally, MS 3, 7-Hydroxyaloin B, and 10-Hydroxyaloin-A demonstrated substantial binding affinity with scores of -9.57, -7.07, and -8.57 kcal/mole against the Spike, RdRP, and Mpro targets, respectively. MD showed the stability and energetically favorable complex formation ability of Isoquercetin and 10-Hydroxyaloin-A. An assessment of drug likeness and physiochemical properties for the test compounds yielded satisfactory results. Various parameters, including RMSD, RMSF, Rg, SASA, hydrogen bond formation, energy landscape, and principal component analysis, collectively indicated the formation of stable and energetically favorable complexes between the lead phytochemicals and the target proteins. Furthermore, an assessment of drug likeness and physiochemical properties for the test compounds yielded satisfactory results.

Enmozhi *et al,* (2021) .[296] performed molecular docking and ADMET properties studies on and rographolide. In the molecular docking analysis, it was observed that andrographolide exhibited a superior docking score compared to the proposed inhibitor PRD_002214 (free energy value of -3.094357 KJ/mol) within the grid box. The compound engaged in interactions, including hydrogen bonds and salt bridges, within the active binding pocket. Notably, the compound established binding interactions with the protease through the formation of four hydrogen bonds involving three distinct residues: Gly143, Cys145, and Glu166. The ADMET properties investigation showed that it exhibited high gastrointestinal absorption, did not permeate the blood-brain barrier, served as a substrate for P-glycoprotein, and did not inhibit specific cytochromes. Notably, andrographolide adhered to drug-like properties, satisfying both Lipinski's Rules and various drug-likeness score criteria, including those defined by Ghose, Veber, Egan, and Muegge.

Khan *et al,* (2021) .[297] employed molecular docking and molecular dynamic simulations to target the chymotrypsin-like protease (3CLpro). Their research unveiled promising inhibitors of this enzyme, including three FDA-approved drugs (remdesivir, saquinavir, and darunavir) and two natural compounds (flavone and coumarin derivatives).

Gajjar *et al,* (2021) .[298] studied 177 natural antiviral compounds and their derivatives using molecular docking, MD simulations and ADMET properties . Molecular docking revealed tannic acid against RdRp (-16.5kcal/mol) and rutin against Mpro (-11.8 kcal/mol). Subsequent molecular dynamics simulations assessed the stability of these ligand-protein complexes. ADMET analysis and Lipinski filters were applied to assess their drug-likeness.

Rudrapal *et al,* (2022) .[299] found that six phytochemicals as promising anti-SARS-CoV-2 agents with notable inhibitory effects against protease (Mpro and PLpro) and polymerase (RdRp) enzymes. These compounds, namely ginsenoside Rg2, saikosaponin A, somniferine, betulinic acid, soyasapogenol C, and azadirachtin A, exhibited strong binding affinities and stable interactions with their respective target proteins. Among them, ginsenoside Rg2, saikosaponin A, and somniferine stood out as the most potent in silico inhibitors, showing potential against Mpro, PLpro, and RdRp, respectively in molecular docking, MD simulations and ADMET profiles.

Kawsar *et al,* (2022) .[300] investigated the binding affinities and interactions between cytidine derivatives and SARS-CoV-2 RdRp through molecular docking analysis, MD

simulation and ADME. The modified derivatives displayed strong interactions with key residues, specifically Pro620 and Lys621. Molecular dynamics simulations lasting 200 ns confirmed the stable docking of these compounds within the RdRp binding pocket. Notably, the derivatives consistently exhibited strong binding with RdRp throughout the simulations, indicative of enhanced binding affinity. The intermolecular interactions were governed by both Van der Waals and electrostatic energies. Furthermore, an ADME evaluation of the optimized inhibitors affirmed their safety profile, indicating improved kinetic properties. Consequently, these selected cytidine derivatives show promise as potential inhibitors against SARS-CoV-2.

Belhassan *et al* (2022) .[301] conducted a molecular docking analysis involving 28 vitamins and their derivatives to evaluate their interactions with the active site of the SARS-CoV-2 main protease. Among these compounds, Folacin, Riboflavin, and Phylloquinone oxide (Vitamin K1 oxide) displayed the most favorable binding energies, indicating their potential as inhibitors. Molecular Dynamics simulations further assessed the binding stability of these vitamins with the SARS-CoV-2 main protease. Riboflavin exhibited the highest stability, forming numerous hydrogen bonds and consistent protein interactions. In contrast, Phylloquinone oxide and Folacin displayed some instability and fewer interactions. Riboflavin also met Lipinski's rule and demonstrated acceptable ADMET properties, suggesting its potential as an effective antiviral candidate for COVID-19 treatment.

Hanine Hadni *et al,* (2022) .[302] applied molecular docking on a set of 26 natural compounds from the flavonoid family, targeting the 3CLpro and RBD sites of SARS-CoV-2. The results of the docking study identified several potential inhibitors for these sites. Silibinin, Tomentin A, Tomentin B, 4′-O-methyldiplacone, Hesperidin, Amentoflavone, and Bilobetin showed promise as inhibitors of SARS-CoV-2 3CLpro, while Herbacetin, Morin, Silibinin, Tomentin E, Amentoflavone, Bilobetin, Baicalein, and Quercetin displayed potential as inhibitors of SARS-CoV-2 RBD. Notably, three compounds demonstrated inhibitory activity against both sites, suggesting their potential in combating COVID-19. Additionally, the evaluation of ADMET pharmacokinetic properties and in silico drug likeness analysis identified six compounds as potential effective drugs against COVID-19. To confirm the stability of hydrogen bonding interactions with crucial residues, molecular dynamics simulations and MM-GBSA calculations were employed, providing further support for the potential of these compounds in countering SARS-CoV-2.

3. Ligand-based drug design

Ligand-based drug design is a pharmaceutical research approach that focuses on designing new drugs or optimizing existing ones based on the properties and interactions of specific molecules, known as ligands that bind to a target protein or receptor in the body. This approach primarily relies on information about the ligands' structure, binding affinity, and other characteristics to guide drug development. LBDD majorly include pharmacophore modeling, quantitative structure-activity relationships QSARs, and artificial intelligence AI[202].

3.1. Pharmacophore modeling

A pharmacophore model elucidates the spatial arrangement of chemical features within ligands that are essential for interacting with the target receptor[303]. These chemical features encompass hydrogen bond donors, hydrogen bond acceptors, aromatic ring systems, hydrophobic regions, positively charged ionizable groups, and negatively charged ionizable groups[304]. Pharmacophore-based virtual screening can identify ligands with varying molecular scaffolds but sharing a similar spatial arrangement of crucial interacting functional groups.

Incorporating the bioactive conformation of molecules within the target binding site is a common practice in developing pharmacophore models. Furthermore, pharmacophore models are frequently employed in molecular alignment during Quantitative Structure-Activity Relationship (QSAR) studies[305]. Several programs, such as Catalyst, PHASE, LigandScout, GALAHAD, and PharmMapper (as listed in Table II.2), enable the automatic construction of pharmacophore models.

To enhance the flexibility of the model, it is often refined by incorporating spatial constraints in regions occupied by inactive molecules. Features that are inconsistently detected in active molecules may be rendered optional or excluded from the final model[212]. The generated pharmacophore model should strike a balance between sensitivity and specificity to minimize the risk of false negatives and false positives. Validation using an independent external test set is crucial[306]. In cases where information about the 3D structure of a receptor and a set of known active compounds is unavailable, a sequence-derived 3D pharmacophore model can prove valuable. For instance, Pharma3D utilizes knowledge of 3D crystal structures and homology models to derive common sequence motifs critical for receptor-ligand interactions in protein families[306,307]. Pharmacophore modeling along with more available tools has been extensively detailed in literature[306–317].

Table II.2 Tools for pharmacophore modeling

Tool	Description	Reference
Catalyst	Identifies common three-dimensional configurations of chemical features among ligands and scores them based on rarity and prevalence.	[318]
LigandScout	Automated tool for generating pharmacophore models based on protein-ligand interactions, including hydrogen bonds, charge transfers, and lipophilic regions. Used for high-throughput virtual screening.	[319]
PharmaGist	Freely available web server for generating ligand-based pharmacophore models using drug-like molecules with binding affinity to a target protein.	[311]
PharmMapper	Identifies potential target receptors for small molecules using	[312]

	pharmacophore mapping.	
Pharmer	Pharmacophore search program that organizes molecular data for rapid screening of large molecule databases.	[313]

3.2. Quantitative Structure-Activity Relationship QSAR

QSAR studies operate on the fundamental concept that changes in the biological activity of compounds can be linked to alterations in their molecular structures[207]. They find extensive application in drug discovery, particularly during hit-to-lead identification and lead optimization phases. These studies involve constructing a statistical model based on correlations, which can then be employed to predict the biological activity of new molecules[305]. To generate a dependable QSAR model, certain key prerequisites must be met: (i) a sufficient quantity of datasets with biological activity information, acquired through standardized experimental procedures, (ii) careful selection of compounds for the training and test sets, (iii) avoidance of autocorrelation among the physiochemical properties of ligands that could lead to data overfitting, and (iv) validation of the model's applicability and predictability using internal and external validation techniques[320].

Molecular structures are represented by molecular descriptors. Molecular descriptors are numerical or mathematical representations of molecular properties that provide information about the structure, composition, and properties of a molecule[321–323]. These descriptors play a crucial role in computational chemistry, drug discovery, and chemoinformatics, converting complex molecular structures into numerical values to quantitatively characterize molecules. There are various types of molecular descriptors, including topological descriptors that describe connectivity, geometric descriptors that represent shape and size, electrostatic descriptors related to charge distribution, quantum chemical descriptors derived from quantum mechanical calculations, and chemical fingerprints encoding structural features[321–326]. Molecular descriptors find applications in drug discovery, toxicity prediction, property estimation, and structure-activity relationship studies. They are vital for virtual screening, toxicity assessment, property prediction, and machine learning models. The selection of relevant descriptors is essential, often requiring statistical techniques for feature selection and dimensionality reduction[322]. Table II.3 resumes the most commonly used tools and softwares for descriptors calculations.

On the other hand, Molecular fingerprints are a type of molecular representation used in cheminformatics and computational chemistry. They are binary or bit-string representations of chemical structures that encode information about the presence or absence of specific structural features within a molecule. Molecular fingerprints are widely used for tasks such as similarity searching, virtual screening, and machine learning in drug discovery and chemical informatics. Calculations of fingerprints follow these steps:

- Binary Encoding: Each molecular fingerprint is a binary vector, where each element (bit) in the vector represents the presence (1) or absence (0) of a particular structural feature or substructure within the molecule.

- Substructure Patterns: Molecular fingerprints are generated by scanning the chemical structure of a molecule and identifying predefined substructure patterns or fragments. These patterns can be as simple as single atoms or more complex, representing functional groups or ring systems.
- Hashing: To create a fixed-length fingerprint, a hashing function is applied to the identified substructures. This function maps the substructures to specific positions in the fingerprint, setting the corresponding bits to 1.

Molecular fingerprints are used to compare the structural similarity between molecules. The more similar the fingerprints of two molecules, the more structurally alike they are considered to be. They are used for various tasks, including similarity searching (to find molecules similar to a query compound in a chemical database), virtual screening (to predict potential drug candidates by comparing their fingerprints to known active compounds), clustering (to group molecules with similar structures), and machine learning (as input features for machine learning models in predictive modeling tasks).

Common types of molecular fingerprints include:

- MACCS Keys: A set of predefined substructures.
- ECFP (Extended Connectivity Fingerprints): Captures molecular connectivity information.
- PubChem Fingerprint: Based on PubChem's collection of chemical substructures.
- Daylight Fingerprint: Created using SMARTS patterns.

Table II.3 Molecular descriptors and fingerprints calculation tools

Programs	**Molecular Descriptors**	**Reference**
BlueDesc	Predicts more than 174 properties, including 3D descriptors and can read 2D structures, 3D structures and SMILES files.	http://www.scbdd.com/chemdes/
ChemAxon	Offers a wide range of chemical calculations, such as molecular weight, elemental composition, LogP, pKa, LogD, LogS, hydrogen bond donor/acceptor (HBDA) count, and various 2D topological and 3D geometrical descriptors.	https://chemaxon.com/products/calculators-and-predictors
PaDEL-Descriptor	A standalone software for computing molecular descriptors and fingerprints, encompassing 797 descriptors (including 663 1D and 2D descriptors and 134 3D descriptors)	[327]

	and 10 types of fingerprints.	
E-DRAGON	The electronic counterpart of the renowned DRAGON software, equipped with over 1,600 molecular descriptors used to evaluate structure-activity or structure-property relationships.	http://www.vcclab.org/lab/edragon/
DRAGON 7.0	Calculates a comprehensive set of 5,270 molecular descriptors, covering atom types, functional groups, fragment counts, topological and geometrical descriptors, and three-dimensional descriptors organized into thirty logical categories.	https://chm.kode-solutions.net/products_dragon.php
Pre-ADMET	A web-based tool for computing drug-like physicochemical descriptors such as lipophilicity (logP), molecular weight, polar surface area, and water solubility.	https://preadmet.bmdrc.kr/
QikProp	Enables the prediction of various pharmacologically relevant descriptors for chemical compounds, including octanol/water and water/gas log Ps, log S, log BB, overall CNS activity, Caco-2, MDCK cell permeabilities, log KHSA for human serum albumin binding, and more.	https://www.schrodinger.com/qikprop
ACD/labs	A versatile tool for calculating various physicochemical descriptors, including aqueous solubility, boiling point/vapor pressure, logD, logP, pKa, and ADME properties.	https://www.acdlabs.com/
Corina Symphony	A cheminformatic tool designed for structure representation and the computation of six types of molecular descriptors: global molecular descriptors, shape descriptors, quantum-mechanical properties, 2D property-weighted autocorrelation, 3D property-weighted autocorrelation,	https://www.mn-am.com/products/corinasymphony

	property-weighted radial distribution functions (RDF), and autocorrelation of surface properties.	
MOPAC	A semiempirical molecular orbital package used for investigating solid-state and molecular structures and reactions. It provides calculations of various descriptors for molecules, radicals, ions, and polymers, including vibrational spectra, thermodynamic parameters, isotopic substitution effects, and force constants.	http://openmopac.net/
AlvaDesc	alvaDesc is a comprehensive software for molecular descriptor calculation, fingerprints, and property prediction. It includes 3D and physicochemical properties descriptors. It also provides tools for data exploration, structure verification, and variable reduction.	[324]

QSAR can be categorized into six different types based on the descriptors used: (i) 1D-QSAR, which examines the correlation between global molecular properties like logP and pKa with biological activities, (ii) 2D-QSAR, where biological activities are linked to structural patterns such as 2D-pharmacophores and connectivity indices, (iii) 3D-QSAR, which explores how biological activities are related to noncovalent interaction fields around the ligands, (iv) 4D-QSAR, an extension of 3D-QSAR that includes an ensemble of ligand configurations, (v) 5D-QSAR, which incorporates various induced-fit models into 4D-QSAR, and (vi) 6D-QSAR, further extending 5D-QSAR by including different solvation models[322]. Some examples of 3D QSAR programs include the HypoGen module of Catalyst[328], PHASE[329], comparative molecular field analysis (CoMFA), and comparative similarity indices analysis (CoMSIA)[330]. QSAR techniques can be classified into two categories: linear and nonlinear, depending on the chemometric methods used. Linear methods encompass linear regression (LR), multiple linear regression (MLR), partial least squares (PLS), principal component analysis (PCA), and principal component regression (PCR). Examples of nonlinear QSAR methods include k-nearest neighbors (kNN), artificial neural networks (ANN), and Bayesian neural nets[321,322]. Extensive and detailed explanation of QSAR principle, techniques, and validation can be found in literature[43,207,208,211,305,320–322,331–339].

3.3. Artificial Intelligence

Artificial intelligence (AI) constitutes a form of machine intelligence hinged upon computers' capacity to glean knowledge from pre-existing data. AI has found utility in various computational modeling techniques aimed at forecasting the biological effects and potential toxicities of drug compounds[340]. Moreover, AI possesses a broad spectrum of applications in drug discovery, encompassing tasks such as forecasting protein folding, protein-protein interactions, virtual screening, quantitative structure-activity relationship (QSAR) analysis, assessment of pharmacokinetic and pharmacodynamic properties (ADMET), and the creation of new drug designs[341]. In rational drug design, two potent methodologies come to the forefront: machine learning (ML) and deep learning (DL)[342]. Among the ML algorithms that have played pivotal roles in drug discovery are support vector machine (SVM)[203], Random Forest (RF)[343], and Naive Bayesian (NB)[344]. On the DL front, notable techniques include convolutional neural networks (CNN), deep neural networks (DNN), recurrent neural networks (RNN), autoencoders, and restricted Boltzmann machines (RBN)[345].

While traditional QSAR approaches effectively predict simpler physicochemical properties like logP and solubility, they often fall short when it comes to forecasting intricate biological properties such as drug efficacy and potential side effects due to their reliance on modest training datasets[346] and limitations in covering the vast chemical space[347]. Coping with the massive data poses significant challenges for both conventional QSAR techniques and machine learning models[348]. In response, AI methods have emerged to tackle the vast and multi-dimensional nature of this big data, enhancing predictions of drug efficacy and potential side effects in animals or humans[202].

A significant challenge in machine learning techniques lies in the need to fine-tune a considerable number of parameters before initiating the algorithm. The performance of regression algorithms is notably influenced by the values of these parameters, which are typically set by default or manually configured through trial and error[349].

As an alternative, metaheuristic optimization algorithms offer the potential to address parameter tuning and overfitting issues in machine learning techniques due to their guided search for near-optimal solutions within minimal computational time[350]. Metaheuristic optimization algorithms are a class of computational techniques used to solve complex optimization problems. These algorithms are particularly useful when dealing with problems that involve searching through large solution spaces, have no known analytical solutions, or are computationally expensive to solve exhaustively. Metaheuristic algorithms are inspired by natural processes, social behaviors, or other abstract concepts and are designed to explore and exploit solution spaces efficiently[351]. Examples are Dragonfly Algorithm DA[352], Particle Swarm Optimization PSO[353], Ant Colony Optimization ACO[354], Simulated Annealing SA[355], Grey Wolf Optimizer GWO[356], and Whale Optimization Algorithm WOA[357]…They work by iteratively improving potential solutions to find near-optimal or optimal solutions. Unlike traditional optimization methods that rely on mathematical properties or derivatives, metaheuristic algorithms focus on heuristic approaches to problem-solving[351].

Specifically, the Dragonfly algorithm DA, a relatively recent addition to swarm optimization techniques[358], stands out for its preference over other algorithms. It is less prone to getting stuck in local minima, which enhances precision and reduces the likelihood of overfitting. Additionally, it significantly reduces computation time[359]. Moreover, particle swarm optimization PSO has demonstrated its effectiveness in enhancing the robustness and performance of machine learning algorithms compared to other optimization methods like the genetic algorithm GA[360]. PSO offers advantages such as ease of implementation, fewer parameters to configure, and superior accuracy with fewer iterations[361].

However, it's worth noting that while AI holds immense potential for efficiently identifying preclinical drug candidates, accurately predicting the binding affinity between a drug molecule and a receptor using AI remains a formidable task, primarily due to several factors. Firstly, AI operates as a data-driven technique, heavily reliant on the quantity and quality of available data[202]. The variability in data sources, especially those stemming from diverse biological assays, and the dearth of high-quality data in public databases present challenges to effective AI learning[362,363].

4. LBDD literature review

Yu (2021) .[364] developed a QSAR model to forecast the inhibitory constant pKi for a set of 89 compounds against 3CLpro. The model was constructed using support vector machine and genetic algorithm techniques. Six specific descriptors were carefully chosen, and the optimal hyperparameters were determined as $C = 90.234$ and $\gamma = 1.19810–5$. The model yielded coefficients of determination R^2 of 0.839 for the training set and 0.747 for the test set. Additionally, the root mean square error (RMSE) was found to be 0.435 for the training set and 0.525 for the test set, respectively.

Janairo *et al*, (2021) .[365] constructed and compared various models to predict the inhibitory activity of the 3CLpro enzyme. These models included SVR, Classification and Regression Trees (CART), random forest, and Multiple Linear Regression (MLR). Among these models, MLR emerged as the top-performing one, achieving notable results with an R^2 value of 0.806 and an RMSE of 0.430 for the test dataset.

Ishola *et al*, (2021) .[366] employed Multiple Linear Regression (MLR) to establish a QSAR model for forecasting the inhibitory activity of known compounds sourced from ChEMBL, specifically targeting the viral 3CLpro enzyme. The resulting model exhibited impressive performance metrics, including a correlation coefficient R of 0.907 and a cross-validation coefficient Q^2 of 0.866.

De *et al*, (2020) .[120] developed a hybrid QSAR model using partial least square and multiple linear regression PLS-MLR on a set of 3CLpro inhibitors. The results showed an R^2 of 0.756 and Q^2 of 0.708. The model was further sustained by molecular docking and MD simulation.

Hammoudan *et al*, (2021) .[367] explored the relationship between the anti-MERS-CoV activities of the HKU4 derived peptides for some peptidomimetic compounds and various

descriptors. The descriptors used were computed using ChemSketch, Marvin Sketch, and ChemOffice software. The principal components analysis (PCA) and the multiple linear regression (MLR) methods were employed. The original data set of 41 peptidomimetic derivatives was randomly divided into training and test sets of 34 and 7 compounds, respectively. The predictive ability of the best MLR model was assessed by determination coefficient R^2 of 0.691, cross-validation parameter Q^2 of 0.528, and the external validation parameter R^2_{test} of 0.794.

Mozafari *et al,* (2021) .[368] combined the adaptive least absolute shrinkage and selection operator (ALASSO) for variable selection, known for its advantageous features, with a nonlinear artificial neural network (ANN) modeling technique. Following a thorough assessment of the ALASSO-ANN model's accuracy and validity (R^2 of 0.86 and RMSE of 0.33), the study utilized relevant descriptors to suggest new compounds and predict their biological activity.

Toropov *et al,* (2022) .[369] conducted a QSAR study using Monte Carlo technique to predict the half maximal inhibitory concentration IC_{50} of aromatic disulfide compounds against 3CLpro. The results showed an R^2 of 0.916 and a Q^2 of 0.681.

Elekofehinti *et al,* (2021) .[370] used an integrated computational approach to identify natural lead compounds. The QSAR model built from Kernel PLS KPLS analysis with radial as binary fingerprint was used to predict the compounds inhibition properties against SARS-CoV-2 3C-like protease. The model had an R^2 of 0.818 and an RMSE of 0.568.

Adhikari *et al,* (2022) .[371] conducted an integrated approach in which multi QSAR models were developed, and validated by molecular docking and MD simulation. A total of 69 diverse chemical compounds, showing potential as SARS-CoV-2 3CLpro inhibitors, were considered for constructing various QSAR models: linear approaches (SW-MLR and GA-MLR), non-linear methods (ANN and SVM), and Hologram QSAR HQSAR techniques. The analysis identified specific structural and physicochemical characteristics essential for inhibiting SARS-CoV-2 3CLpro, notably the presence of 2-oxopyrrolidinyl methyl and benzylester functionalities, as well as the inclusion of a methylene (hydroxy) sulphonic acid warhead group. HQSAR showed the best performance with R^2 of 0.867 and a Q^2 of 0.67.

Chedadi *et al,* (2021) .[372] employed 3D-QSAR, molecular docking, and ADMET prediction, focusing on a series of inhibitors targeting SARS-CoV 3CLpro. The 3D-QSAR investigation specifically utilized the Comparative Molecular Field Analysis (CoMFA) and Comparative Molecular Similarity Indices Analysis (CoMSIA) methodologies. These analyses yielded robust results, with cross-validation coefficient Q^2 values of 0.80. The model's predictive capacity was further supported by determination coefficient R^2_{test} values of 0.725 when evaluated with a test set comprising seven molecules.

Edache *et al,* (2022) .[267] conducted a comprehensive mathematical approach encompassing quantitative structure-activity relationship (QSAR), homology modeling, docking simulations, ADMET analysis, and molecular dynamics simulations on iminoguanidine derivatives. The QSAR method was employed to statistically examine the

structure-activity relationships (SAR) and yielded highly significant results for a robust predictive model (GA-MLR: Q^2_{LOO} = 0.8027; R^2 = 0.8735; R^2_{test} = 0.7536). In-depth analysis of these predictive models revealed that the Centered Broto-Moreau autocorrelation at lag 1, weighted by I-state, and the 3D topological distance-based autocorrelation at lag 9, weighted by I-state, played pivotal roles in governing the biological activity. These descriptors provided valuable insights into the properties necessary for the development of potent inhibitors against Pseudomonas aeruginosa.

Gholivand *et al* (2022) .[373] assessed a set of 51 compounds (consisting of various mono-, bis-, and tetra-phosphonates designed as derivatives of Remdesivir, among which, 32 were newly developed and thoroughly characterized) through a combination of molecular docking, QSAR, and pharmacophore modeling. Their biological activities were evaluated. The findings from the QSAR analysis revealed that the polarity and molecular topology of these phosphonate derivatives played pivotal roles in influencing their binding energy and inhibitory potential. The Genetic Algorithm-A stepwise multiple linear regression (GA-SW-MLR) model demonstrated strong predictive capability, achieving an R^2 value of 0.843 and an RMSE of 0.341.

Khanfar (2021) .[374] employed a combination of structure-based pharmacophore modeling and validated QSAR analysis to identify novel IRAK-4 inhibitors from a natural products database. The MLR-QSAR model, with an R^2 value of 0.78, integrated molecular descriptors with a structure-based pharmacophore. This approach effectively explained the variation in bioactivity among a diverse set of IRAK-4 inhibitors. In experimental in vitro testing, the top-ranked compounds: uvaretin, saucerneol, and salvianolic acid B were confirmed as active IRAK-4 inhibitors, demonstrating IC_{50} values in the low micromolar range.

Saha *et al,* (2023) .[375] 2D MLR-QSAR model developed 84 anti-SARS-CoV2 compounds gathered from the literature, and descriptors were calculated using PADEL software. The model showed a performance of with R^2 and Q^2 values of 0.87028 and 0.70493, respectively. The findings also indicate that the electronic features and E-state of atoms in the molecules are directly linked to their anti-SARS-CoV-2 drug activity.

Soleymani *et al,* (2023) .[376] created validated QSAR models using CORAL software and Monte Carlo optimization to predict the inhibitory activity of 81 compounds based on isatin and indole against SARS CoV 3CLpro. These models employ the index of ideality correlation (IIC) as an objective function optimization method. Optimal descriptors were chosen from the hybrid model, which combines SMILES and hydrogen-suppressed graph (HSG) based on the objective function. The model had an R^2 of 0.96 and Q^2 of 0.99. Through model interpretation, eight synthesized compounds were identified and sourced from the ChEMBL database, displaying potential as effective inhibitors of SARS CoV 3CLpro. Docking and ADMET studies revealed that compounds CHEMBL4458417 and CHEMBL4565907, both featuring an indole scaffold, exhibit positive drug-likeness values and the highest drug-score, making them promising lead candidates.

Abdul Amin *et al,* (2021) .[377] developed a Monte Carlo optimization-based QSAR model, applying it to virtually screen a selection of in-house chemicals. SwissADME's in silico

analysis identified thirteen compounds with favorable drug-likeness characteristics. Subsequent molecular docking studies revealed these compounds to exhibit advantageous interactions with the target SARS-CoV-2 Papain Like protease PLpro enzyme, suggesting their potential as a foundation for designing and optimizing drugs against SARS-CoV-2 PLpro.

Ghosh *et al,* (2021) .[378] built a Monte Carlo optimization-based QSAR model to screen a library of natural product hits. Analysis of the fragments within these active molecules hinted at the possibility of synthesizing novel inhibitors for the SARS-CoV-2 Mpro enzyme by combining fragments or attaching them to other structural scaffolds.

Law *et al,* (2020) .[379] generated a ligand-based pharmacophore model using established antiviral drugs. This model was employed to assess the antiviral potential of twenty vanillin derivatives as inhibitors of SARS-CoV-2's Mpro. Furthermore, the structure-based pharmacophore model indicated that vanillin derivatives (1-20) exhibited highly promising attributes, suggesting their potency as antiviral compounds against COVID-19.

Rafi *et al,* (2020) .[380] selected lopinavir and compounds with structural similarities for binding to the main protease, while favipiravir was chosen as a candidate for targeting the RNA-dependent RNA polymerase. Through molecular docking and a quantitative structure-activity relationships (QSAR) analysis, it was discovered that these chosen compounds exhibited favorable binding affinities. However, they exhibited limitations in terms of their drug-like properties.

Thorat *et al,* (2023) .[381] elaborated 2D and 3D QSAR models using a dataset of 92 2-aryl-4-quinoline carboxylic acid analogs known as DHODH inhibitors. The correlation coefficients R^2 for the training set of Partial Least Squares PLS and all five Kernel Partial Least Squares (KPLS) models for their respective fingerprints are as follows: 0.7091 (linear), 0.8336 (linear), 0.7586 (radial), 0.8606 (dendritic), 0.6832 (desc), and 0.7670 (molprint2D). These values indicate strong correlations ($R2 \approx 0.9$ but > 0.6).

Shaker *et al,* (2020) .[205] applied Rosalind platform for the rapid discovery of SARS CoV2 replication inhibitors that target the virus main protease Mpro. Through a series of training and optimization rounds based on reported SARS CoV2 Mpro inhibitors helped by docking into the recently reported crystal structures of SARS CoV2 Mpro and medicinal chemistry input, a series of promising SARS CoV2 Mpro inhibitors were identified.

5. Virtual screening

Virtual screening is a computational method employed in drug discovery to screen large collections of compounds for potential drug candidates. This technique uses computer algorithms to predict the ability of small molecules to bind to a target protein[382]. This allows to efficiently select promising compounds for subsequent evaluation, saving and resources compared to conventional experimental screening methods. Virtual screening encompasses both ligand-based and structure-based approaches and includes a range of computational tools, including docking, similarity analysis, pharmacophore-based searches,

and machine learning algorithms[382]. The objective of virtual screening is to identify promising candidates with powerful properties that can undergo further refinement into potential drugs, all the while reducing the time and resources required for experimental screening[382]. Virtual screening was extensively explained in literature[382–387].

Virtual screening can be both a structure-based technique and a ligand-based technique, depending on how it is implemented and the type of data used.

5.1. Structure-Based Virtual Screening (SBVS)

- In SBVS, the primary focus is on the 3D structure of the target protein. It involves docking small molecules into the binding site of the protein and evaluating their binding affinity.

- SBVS utilizes information about the protein's structure, such as the coordinates of its atoms and the shape and electrostatic properties of the binding site.

- It is considered a structure-based technique because it relies on the structural information of the target protein[382].

5.2. Ligand-Based Virtual Screening (LBVS)

- In LBVS, the primary focus is on the chemical properties and activities of known ligands or compounds that have interacted with the target.

- LBVS relies on the similarity between these known ligands and the compounds in a screening library.

- It does not directly use the 3D structure of the target protein but is based on the assumption that compounds similar to known active ligands will also interact with the target.

- LBVS is considered a ligand-based technique because it relies on the chemical characteristics of the ligands without direct consideration of the protein's 3D structure[382].

In practice, researchers often use a combination of both SBVS and LBVS in virtual screening workflows to increase the chances of identifying potential drug candidates. SBVS is well-suited for finding compounds that interact with specific binding sites on target proteins, while LBVS can identify molecules with similar properties to known active ligands. This hybrid approach takes advantage of both structural and chemical information to improve the efficiency of virtual screening in drug discovery[382].

6. Conclusion

In the realm of pharmaceutical research, Computer-Aided Drug Design CADD has emerged as a transformative force, reshaping the landscape of drug discovery and development. This chapter provided a comprehensive overview of three fundamental components of CADD: Structure-Based Drug Design SBDD, Ligand-Based Drug Design LBDD, and Virtual Screening. These methodologies have demonstrated the ability to accelerate the identification of potential drug candidates, optimizing molecular interactions, and streamlining the drug development process. The following chapter will detail the modeling part including the

methodology, the algorithms, and the softwares used as well as the results obtained and their extensive discussion.

CHAPTER III
Modelling Procedure

1. Objectives

The Methodology chapter of this thesis encompasses an intricate integration of two distinct yet complementary Computer-Aided Drug Design (CADD) methodologies aimed at the identification of prospective inhibitors against SARS-CoV-2. The study extensively employs both Structure-Based and Ligand-Based approaches to navigate the complex landscape of drug discovery.

Within the ambit of Structure-Based methodologies, emphasis is placed on scrutinizing Absorption, Distribution, Metabolism, Excretion, and Toxicity (ADMET) properties alongside molecular docking analyses. These approaches offer a detailed comprehension of the molecular interactions and essential structural characteristics of potential inhibitors. Conversely, the Ligand-Based methodologies delve into the intricate realm of Quantitative Structure-Activity Relationship (QSAR) modeling, employing a diverse range of algorithms such as Support Vector Regression (SVR), Particle Swarm Optimization-Support Vector Regression (PSO-SVR), Dragonfly Algorithm-Convolutional Neural Network (DA-CNN), Multilinear Regression (MLR), and Bayesian Regulation Backpropagation Neural Network (BR-BPNN).

The delineation of the Methodology chapter is structured around two distinct datasets. The first dataset encompasses ML300 analogues inhibitors, where the application of QSAR models involving MLR, SVR, and PSO-SVR has been meticulously executed. The second dataset, a heterogeneous collection, involves the implementation of BR-BPNN, DA-CNN, and DA-SVR QSAR methodologies, alongside ADMET property analysis and molecular docking studies.

The overarching objective of this multifaceted methodological approach is to discern potential inhibitors of SARS-CoV-2. Through the comprehensive exploration and analysis of diverse datasets using an array of sophisticated CADD tools, this research aims to contribute substantively to the ongoing quest for novel therapeutic interventions against this virulent pathogen.

2. Methodology

2.1. ML300 analogues

The methodology used is illustrated in Fig III.1. QSAR models: MLR, SVR, and PSO-SVR were developed.

2.1.1. Data set retrieval

The logic behind the selected nonpeptidic and noncovalent molecules is their proven activity against 3CLpro SARS-CoV-2 and their novelty[388]. The selected compounds are derivatives of ML300 (Fig. III.2.a), a potent probe noncovalent inhibitor of the SARS-CoV-1 3CLpro developed by Molecular Libraries Probe Production Centers Network (MLPCN), that despite its nanomolar activity, it was far from being optimized and suffered from poor metabolic stability. A novel benzotriazole-based derivatives series was synthesized. They have a much

more robust nanomolar and submicromolar inhibitory potential in live SARS-CoV-2, with a better metabolic profile[388]. These derivatives were intriguing for several reasons. Their

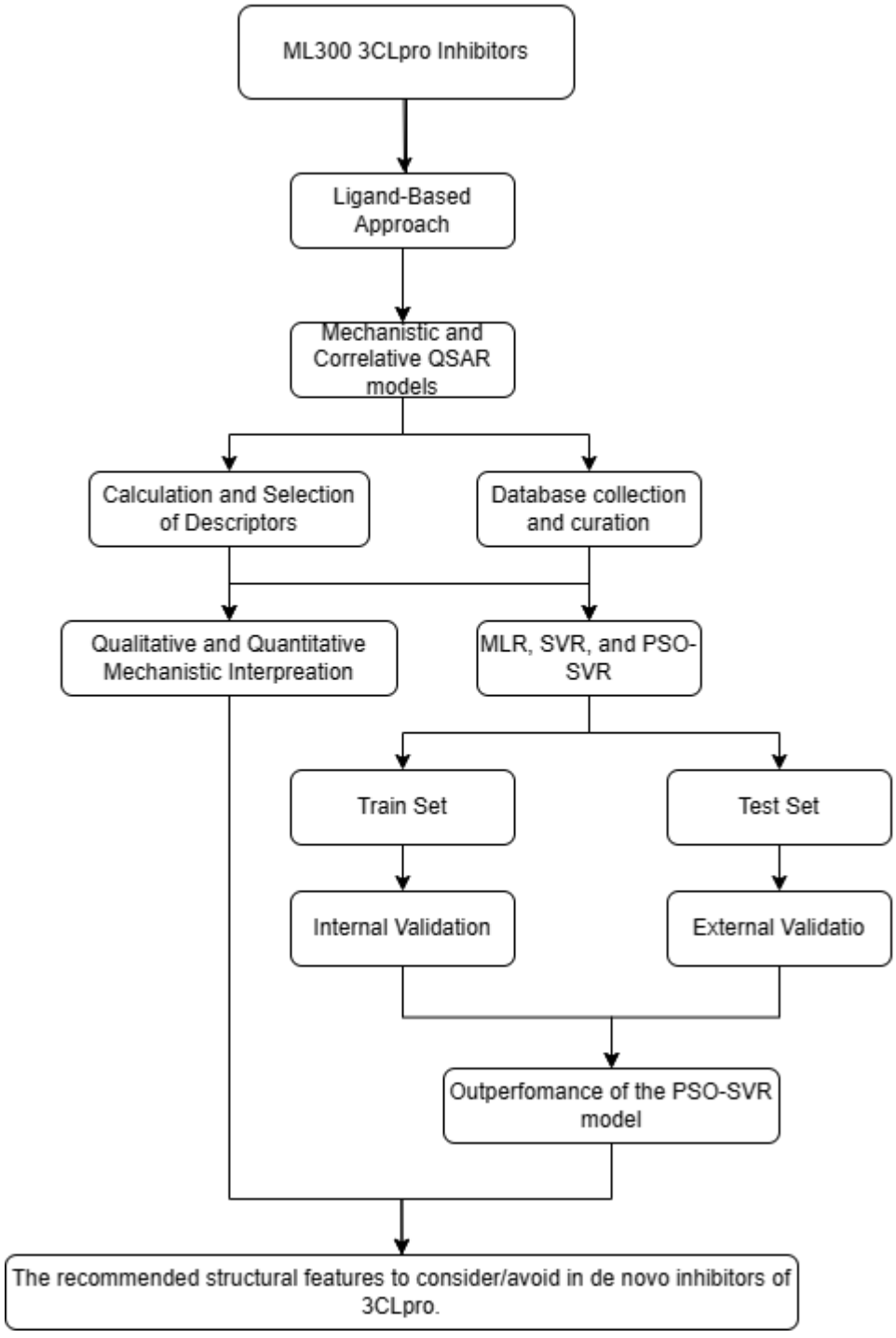
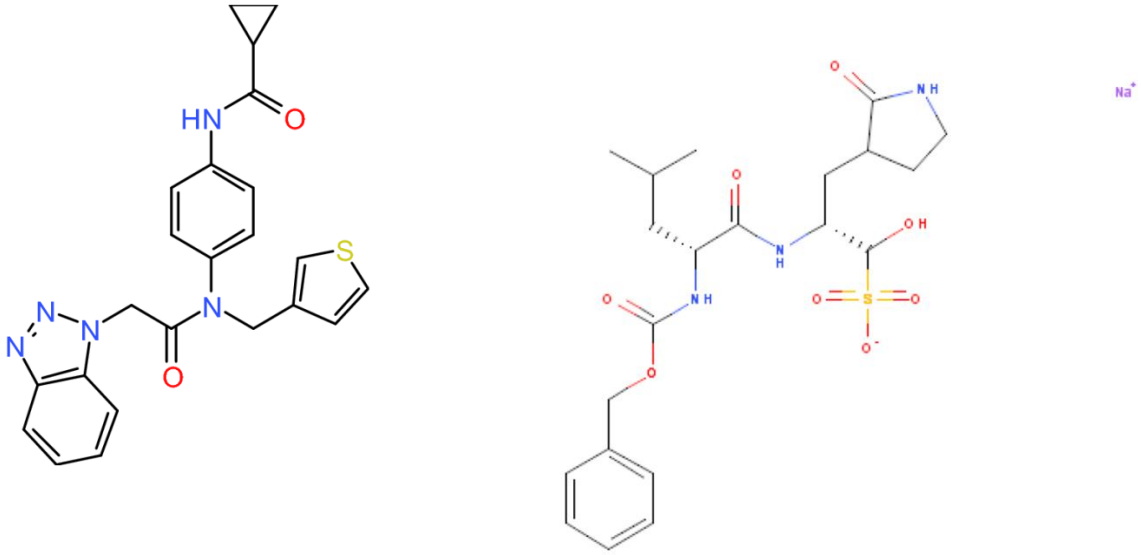


Fig III.1 organizational chart of the ML300 analogues modelling

experimental activity values were determined, and thus the interpretation of a QSAR model is applicable. Otherwise, data may risk being heterogeneous (same molecule but with two or

more activity values from different resources). Also, their mechanism of action is well-determined and different from the recently proven inhibitor GC376 (Fig. III.2.b). These molecules follow a noncovalent binding mode, while the GC376 follows a covalent binding mode with the residues in the enzyme's active site. The original articles[388] and [389] explain the detailed mechanisms for the ML300 derivatives series and GC376, respectively. It is worth mentioning that, generally, noncovalent inhibitors are advantageous over covalent inhibitors in terms of low toxicity and broader inhibition spectrum[390]. This means that noncovalent inhibitors are not only safer and more likely to reach clinical trials and approval, but they can also affect similar targets, which means they are more likely to be repurposed or repositioned. For example, ML300 analogues can target HKU4, a suspected reservoir host for MERS[391]. Finally, the quantitative structure–activity for these derivatives remains unknown[388], making modelling this relationship necessary. Thus, the QSAR models built on the selected database open the door to discovering novel potent noncovalent inhibitors, different from the proven covalent ones, and pave the way for further docking, virtual screening, and synthesis research to identify potential hits based on its findings.



(a) *(b)*

Fig.III.2 2D structures of noncovalent and covalent inhibitors: (a) ML300,(b) GC376

The visualization of the molecules' structures and the generation of their 3D structures with minimal energy were carried out using the Data-Warrior software[392]. All the used molecules are presented in Table S-III.1, along with their IC_{50} values and other information. For linearization purposes, the experimental values of the half maximal inhibitory concentration (IC_{50}, nM) were converted to 9 minus the negative decimal logarithmic scale ($pIC_{50} = 9\text{-log}\ (IC_{50})$).

Table III.1 ML300 analogues database

SMILES Notation	Structure No	IC50 (nM)	Institution
Clc1cc(ccc1)CN(C(=O)Cn1nnc2c1cccc2)c1ccc(cc1)C1=NCN=C1	1	19	Cleveland Clinic
Clc1cc(cc(F)c1)CN(C(=O)Cn1nnc2c1cccc2)c1ccc(cc1)-c1cccnc1	4	74	Cleveland Clinic
s1cc(cc1)CN(C(=O)Cn1nnc2c1cccc2)c1ccc(cc1)-c1cn[nH]c1	5	110	Cleveland Clinic
Clc1cc(cc(Cl)c1)CN(C(=O)Cn1nnc2c1cccc2)c1ccc(cc1)-c1cccnc1	7	111	Cleveland Clinic
s1cc(cc1)CN(C(=O)Cn1nnc2c1cccc2)c1ccc(cc1)C1=NCN=C1	9	150	Cleveland Clinic
Clc1cc(ccc1)CN(C(=O)Cn1nnc2c1cccc2)c1cc(OC)c(cc1)-c1cn[nH]c1	13	206	Cleveland Clinic
Clc1cc(ccc1)CN(C(=O)Cn1nnc2c1cccc2)c1ccc(cc1OC)-c1cn[nH]c1	14	208	Cleveland Clinic
Clc1cc(ccc1)CN(C(=O)Cn1nnc2c1cccc2)c1ncc(cc1)-c1cn[nH]c1	15	214	Cleveland Clinic
Clc1cc(ccc1)CN(C(=O)Cn1nnc2c1cccc2)c1ccc(cc1)-c1cccnc1	16	250	Cleveland Clinic
Clc1cc(ccc1)CN(C(=O)Cn1nnc2c1cccc2)c1ccc(cc1)-c1cn[nH]c1	21	334	Cleveland Clinic
Clc1cc(ccc1)CN(C(=O)Cn1nnc2c1cccc2)c1ccc(nc1)-c1cn[nH]c1	22	336	Cleveland Clinic
Clc1cc(cnc1)CN(C(=O)Cn1nnc2c1cccc2)c1ccc(cc1)-c1cccnc1	23	370	Cleveland Clinic
s1cc(cc1)CN(C(=O)Cn1nnc2c1cccc2)c1ccc(cc1)C(=O)N	24	380	Cleveland Clinic
Clc1cc(ccc1)CN(C(=O)Cn1nnc2c1cccc2)c1ccc(cc1)C1=CC=CNC1=O	29	657	Cleveland Clinic
O=C(N(CC1CCC1)c1ccc(cc1)-c1cccnc1)Cn1nnc2c1cccc2	31	780	Cleveland Clinic
s1cc(cc1)CN(C(=O)Cn1nnc2c1cccc2)c1ccc(cc1)C1=CC=CNC1=O	32	780	Cleveland Clinic
Clc1cc(ccc1)CN(C(=O)Cc1cccnc1)c1ccc(cc1)-c1cn[nH]c1	34	864	Cleveland Clinic
s1cc(cc1)CN(C(=O)Cn1nnc2c1cccc2)c1ccc(cc1)-c1ncccc1	35	900	Cleveland Clinic
s1cc(cc1)CN(C(=O)Cn1nnc2c1cccc2)c1ccc(cc1)-c1cccnc1	37	930	Cleveland Clinic
O=C(N(CC1CCCC1)c1ccc(cc1)-c1cccnc1)Cn1nnc2c1cccc2	39	1000	Cleveland Clinic
s1cc(nc1)CN(C(=O)Cn1nnc2c1cccc2)c1ccc(cc1)-c1cccnc1	40	1130	Cleveland Clinic
s1cc(cc1)CN(C(=O)Cn1nnc2c1cccc2)c1ccc(NS(=O)(=O)C)cc1	43	1400	Cleveland Clinic
s1cc(cc1)CN(C(=O)Cn1nnc2c1cccc2)c1ccc(cc1)-c1ccncc1	45	1570	Cleveland Clinic
O=C(N(CC1CC1)c1ccc(cc1)-c1cccnc1)Cn1nnc2c1cccc2	49	2440	Cleveland Clinic
Clc1cc(ccc1)CN(C(=O)Cc1cncnc1)c1ccc(cc1)-c1cn[nH]c1	51	2723	Cleveland Clinic
s1cc(cc1)CN(C(=O)Cn1nnc2c1cccc2)c1ccc(cc1)-c1cccnc1OC	52	3090	Cleveland Clinic
O=C(N(CCC(C)C)c1ccc(cc1)-c1cccnc1)Cn1nnc2c1cccc2	54	4210	Cleveland Clinic
s1cc(cc1)CN(C(=O)Cn1nnc2c1cccc2)c1ccc(NC(=O)C2CC2)cc1	55	4450	Cleveland Clinic

Clc1cc(ccc1F)CN(C(=O)Cn1nnc2c1cccc2)c1ccc(cc1)-c1cccnc1	57	4870	Cleveland Clinic
s1cc(cc1)CN(C(=O)Cn1nnc2c1cccc2)c1ccc(NS(=O)(=O)C2CC2)cc1	60	7980	Cleveland Clinic
s1cc(cc1)CN(C(=O)Cn1nnc2c1cccc2)c1ccc(cc1)-c1ncn(c1)C	64	11310	Cleveland Clinic
s1cc(cc1)CN(C(=O)Cn1nnc2c1cccc2)c1ccc(cc1)C(=O)NC1CC1	66	16550	Cleveland Clinic
s1cc(cc1)CN(C(=O)Cn1nnc2c1cccc2)c1ccc(cc1)-c1ccccc1	67	16820	Cleveland Clinic
s1cc(cc1)CN(C(=O)Cn1nnc2c1cccc2)c1ccccc1	69	19730	Cleveland Clinic
O1C[C@H](CC1)CN(C(=O)Cn1nnc2c1cccc2)c1ccc(cc1)-c1cccnc1	71	20500	Cleveland Clinic
s1cc(cc1)CN(C(=O)Cn1nnc2c1cccc2)c1ccc(cc1)-c1cn(nc1)C	72	22780	Cleveland Clinic
s1cc(cc1)CN(C(=O)Cn1nnc2c1cccc2)c1ccc(cc1)C1=NCN=C1	2	60	Cleveland Clinic
Clc1cc(ccc1)CN(C(=O)Cn1nnc2c1cccc2)c1ccc(cc1)C1=NCN=C1	3	68	Cleveland Clinic
s1cc(cc1)CN(C(=O)Cn1nnc2c1cccc2)c1ccc(cc1)-c1cn[nH]c1	6	110	Cleveland Clinic
Clc1cc(cc(F)c1)CN(C(=O)Cn1nnc2c1cccc2)c1ccc(cc1)-c1cccnc1	8	119	Cleveland Clinic
Clc1cc(ccc1)CN(C(=O)Cn1nnc2c1cccc2)c1ncc(cc1)-c1cn[nH]c1	10	171	Cleveland Clinic
Clc1cc(cc(Cl)c1)CN(C(=O)Cn1nnc2c1cccc2)c1ccc(cc1)-c1cccnc1	11	176	Cleveland Clinic
Clc1cc(ccc1)CN(C(=O)Cn1nnc2c1cccc2)c1ccc(cc1OC)-c1cn[nH]c1	12	197	Cleveland Clinic
Clc1cc(ccc1)CN(C(=O)Cn1nnc2c1cccc2)c1ccc(cc1)-c1cn[nH]c1	17	270	Cleveland Clinic
Clc1cc(cc(c1)C)CN(C(=O)Cn1nnc2c1cccc2)c1ccc(cc1)-c1cccnc1	18	312	Cleveland Clinic
Clc1cc(ccc1)CN(C(=O)Cn1nnc2c1cccc2)c1cc(OC)c(cc1)-c1cn[nH]c1	19	328	Cleveland Clinic
Clc1cc(ccc1)CN(C(=O)Cn1nnc2c1cccc2)c1ccc(nc1)-c1cn[nH]c1	20	333	Cleveland Clinic
Clc1cc(ccc1)CN(C(=O)Cn1nnc2c1cccc2)c1ccc(cc1)-c1cccnc1	25	390	Cleveland Clinic
s1cc(cc1)CN(C(=O)Cn1nnc2c1cccc2)c1ccc(cc1)C(=O)N	27	440	Cleveland Clinic
Clc1cc(cnc1)CN(C(=O)Cn1nnc2c1cccc2)c1ccc(cc1)-c1cccnc1	28	540	Cleveland Clinic
Clc1cc(ccc1)CN(C(=O)Cn1nnc2c1cccc2)c1ccc(cc1)C1=CC=CNC1=O	30	746	Cleveland Clinic
O=C(N(CC1CCC1)c1ccc(cc1)-c1cccnc1)Cn1nnc2c1cccc2	33	830	Cleveland Clinic
s1cc(cc1)CN(C(=O)Cn1nnc2c1cccc2)c1ccc(cc1)C1=CC=CNC1=O	36	920	Cleveland Clinic
s1cc(cc1)CN(C(=O)Cn1nnc2c1cccc2)c1ccc(cc1)-c1cccnc1	38	950	Cleveland Clinic
O=C(N(CC1CCCC1)c1ccc(cc1)-c1cccnc1)Cn1nnc2c1cccc2	41	1160	Cleveland Clinic

Clc1cc(ccc1)CN(C(=O)Cc1cccnc1)c1ccc(cc1)-c1cn[nH]c1	42	1270	Cleveland Clinic
s1cc(cc1)CN(C(=O)Cn1nnc2c1cccc2)c1ccc(cc1)-c1ncccc1	44	1400	Cleveland Clinic
s1cc(nc1)CN(C(=O)Cn1nnc2c1cccc2)c1ccc(cc1)-c1cccnc1	46	1590	Cleveland Clinic
s1cc(cc1)CN(C(=O)Cn1nnc2c1cccc2)c1ccc(NS(=O)(=O)C)cc1	47	1610	Cleveland Clinic
s1cc(cc1)CN(C(=O)Cn1nnc2c1cccc2)c1ccc(cc1)-c1ccncc1	48	2090	Cleveland Clinic
O=C(N(CC1CC1)c1ccc(cc1)-c1cccnc1)Cn1nnc2c1cccc2	50	2510	Cleveland Clinic
Clc1cc(ccc1)CN(C(=O)Cc1cncnc1)c1ccc(cc1)-c1cn[nH]c1	53	3442	Cleveland Clinic
O=C(N(CCC(C)C)c1ccc(cc1)-c1cccnc1)Cn1nnc2c1cccc2	56	4620	Cleveland Clinic
s1cc(cc1)CN(C(=O)Cn1nnc2c1cccc2)c1ccc(NC(=O)C2CC2)cc1	58	4990	Cleveland Clinic
s1cc(cc1)CN(C(=O)Cn1nnc2c1cccc2)c1ccc(cc1)-c1cccnc1OC	59	5070	Cleveland Clinic
Clc1cc(ccc1F)CN(C(=O)Cn1nnc2c1cccc2)c1ccc(cc1)-c1cccnc1	61	8530	Cleveland Clinic
s1cc(cc1)CN(C(=O)Cn1nnc2c1cccc2)c1ccc(NS(=O)(=O)C2CC2)cc1	62	8770	Cleveland Clinic
s1cc(cc1)CN(C(=O)Cn1nnc2c1cccc2)c1ccc(cc1)-c1ncn(c1)C	63	9550	Cleveland Clinic
s1cc(cc1)CN(C(=O)Cn1nnc2c1cccc2)c1ccc(cc1)C(=O)NC1CC1	65	15940	Cleveland Clinic
s1cc(cc1)CN(C(=O)Cn1nnc2c1cccc2)c1ccc(cc1)-c1cn(nc1)C	70	20390	Cleveland Clinic
s1cc(cc1)CN(C(=O)Cn1nnc2c1cccc2)c1ccc(cc1)-c1ccccc1	74	31180	Cleveland Clinic

2.1.2. Descriptors generation and selection

Molecular descriptors are the numerical representation of molecules' structural features; they were calculated by Blue-Desc Descriptor Calculator (http://www.scbdd.com/blue_desc/index/) through the online platform ChemDes[325] to obtain134 3D, 2D, and 1D descriptors of various families (topological, electronegativity, autocorrelation, functional groups). The genetic algorithm and Double Cross-Validation DCV tool[334] were used to select only the representative descriptors. This step removes highly correlated descriptors, null, infinity, constant, and near constant descriptors (minimal deviation of the descriptors' values for the molecules). It leaves only the descriptors that induce a change in the activity value when they change. Hence, the number of descriptors was reduced to 7 impactful descriptors and thus forming a matrix of 71 × 7.

2.1.3. Data division and model development

The data set was optimally distributed into a train set, containing 80% of the dataset, and a test set, including 20% of the dataset. QSAR models were developed using MLR, SVR, and PSO-SVR. The MLR was performed using MLRplus Validation 1.3 software (http://teqip.jdvu.ac.in/QSAR_Tools/) whereas the SVR and PSO-SVR algorithms were conducted using MATLAB environment.

2.1.4. Regression methods

2.1.4.1. Multiple linear regression

Multiple Linear Regression MLR is a type of linear regression that takes more than one explanatory (independent) variable (x_1, …, x_n) to build a multivariate predictive model[393].

The independent variables should not be correlated but uniquely describe the outcome (dependent variable), while their number should be kept below one-fifth of the number of observations. A fast statistical analysis on the descriptors will reveal the over-fitting problem. Accordingly, the number of independent variables can be decreased to avoid over-fitting in the MLR model[394].

2.1.4.2. Support vector regression

Support Vector Regression SVR is a regression technique rooted in Vapnik's concept of support vector. SVR is a kernel-based technique with a solid theoretical basis and is generally easier to implement than artificial neural networks, its major competitor[395].SVR is a supervised-learning machine characterized by using Kernel functions, the number of support vectors, penalty parameters C, coefficient of non-sensitivity ε, and the Gaussian nucleus bandwidth σ[396]. These parameters can be manually defined or by the mean of heuristic optimization algorithms.

2.1.4.3. Particle swarm optimization PSO

Inspired by the behaviour of swarms such as birds' flocks and ant colonies, Particle Swarm optimization is a widespread metaheuristic that Kennedy and Eberhart first developed. It solves complex nonlinear optimization problems with the advantages of having fewer parameters to adjust and an already available literature discussion and community for the parameters that should be adjusted[397]. It consists of finding the solution (global best gbest) by comparing and updating an individual's information of position and velocity (personal best pbest)[398]. PSO is coupled with machine learning algorithms to determine their optimal parameters.

2.1.5. Statistical validation

Statistical validation establishes the reliability and relevance of the built model for the specific target[321]. One of the challenges of validating the QSAR model is to prove it statically accurate in predicting the activity of new molecules unknown to the model[336]. Internal and external validation criteria were calculated using the external validation tool downloaded from https://sites.google.com/site/dtclabxvplus/ and the Leave One Out validation from https://dtclab.webs.com/software-tools. An extensive explanation of the tool and the metrics used has been detailed in the literature[334,367]. Another recent and exciting criterion to compare different models' fit, is Akaike's information criterion AIC. The lowest value of the AIC corresponds to the model that is best fitted to the experimental data[399]. The used statistical validation criteria and their formulas are detailed in Table III.2.

2.1.6. Applicability domain

The Applicability Domain AD is a theoretical space defined by the model's descriptors and the predicted response of the molecules. The model is considered reliable in the AD space.

Thus, the modelled responses outside the AD are considered outliers[322]. The AD estimation approaches can be classified as descriptor-space-based and training-set-response-space-based approaches.

The first includes ranges in the descriptor space; the geometric training-set-response-space-based approach includes the range of the variable response technique. Thus, a molecule is considered outside the AD of a particular model if it satisfies the following conditions: (i) at least one descriptor is outside of the ranges approach, and (ii) the distance between the chemical and the center of the training data set is greater than the acceptable threshold for distance approaches[322]. In the present work, the AD was calculated using the Applicability Domain MATLAB toolbox downloaded from https://michem.unimib.it/.

2.2. Heterogeneous inhibitors

The methodology used to model the heterogeneous dataset consists of combining Structure-Based and Ligand-Based approaches to best identify potential inhibitors. The methodology is illustrated in Fig III.3.

2.2.1. Dataset retrieval and pretreatment

Molecules were extracted from the ChEMBL database. Molecules were visualized and pretreated using the Data Warrior[392] and Alva Molecule (www.alvascience.com/alvaMoecule/) softwares. Molecules with no or a greater value than 10000 nM IC_{50}, the half-maximum inhibitory activity values, were discarded. All the selected molecules have molecular weights between 200 and 800 Da, ensuring that they have a good permeability in the cell membrane, and thus ensuring a good drug likeliness. Duplicated structures were also removed. A sample of the selected inhibitors is shown in Fig III.4, and the extended list of all the molecules along with their corresponding IC_{50} and other information are listed in the Supplementary File as and SDF format in the supplementary material. For linearization purposes, the negative decimal logarithm of the IC_{50} was used.

2.2.2. Molecular descriptors calculation and selection

Molecular descriptors were calculated using AlvaDesc software[324]. A total of 5666 descriptors of all types were obtained. A selection of the representative descriptors was performed using SVM implemented in the Alvamodel software (www.alvascience.com/alvaModel/). The final matrix to be used as an input to the regression algorithm consisted, thus, of 246x10.

2.2.3. Database division

To find the best percentage of the train and test sets, many trials were carried out varying the test set fraction from 0.35 to 0.15. The best results were obtained at the fraction of 0.2. The test and train data were selected randomly.

Table III.2 Validation criteria

Metric	Name	Formula
R^2	Coefficient of determination	$1-\frac{\sum(Y_{obs}-Y_{calc})^2}{\sum(Y_{obs}-\bar{Y}_{obs})^2}$
R_0^2	Determination coefficient without intercept	$1-\frac{\sum(Y_{obs}-k.Y_{calc})^2}{\sum(Y_{obs}-\bar{Y}_{obs})^2}, where: \mathrm{k}=\frac{\sum(Y_{obs}.Y_{calc})}{\sum(Y_{calc})^2}$
Q^2	Leave one out	$1-\frac{\sum(Y_{obs(train)}-Y_{calc(train)})^2}{\sum(Y_{obs(train)}-\bar{Y}_{training})^2}$
$\bar{r}_m^2$	Average r_m^2	$\frac{r_m^2+r_m'^2}{2}, where: r_m^2=r^2\times\left(1-\sqrt{\left(r^2-r_0^2\right)}\right); r_m'^2=r^2\times\left(1-\sqrt{\left(r^2-r_0'^2\right)}\right)$
Δr_m^2	Delta r_m^2	$\left\lvert r_m^2-r_m'^2\right\rvert$
RMSE	Root mean square error	$\sqrt{\frac{\sum(Y_{obs})-Y_{calc})^2}{n}}$
SD	Standard deviation	$\frac{\sum(Y_{obs}-Y_{calc})}{n}$
MAE	Mean absolute error	$\frac{\sum(Y_{obs(train)}-Y_{calc(train)})}{n}$
Q^2_{F2}		$1-\frac{\sum(Y_{obs(test)}-Y_{calc(test)})^2}{\sum(Y_{obs(test)}-\bar{Y}_{test})^2}$
AIC	Akaike's information criterion	$AIC=nln\left(\frac{SSE}{n}\right)+2n_p+\frac{2n_p\left(n_p+1\right)}{n-\left(n_p+1\right)}, where: SSE=\sum(Y_{obs}-Y_{calc})^2$

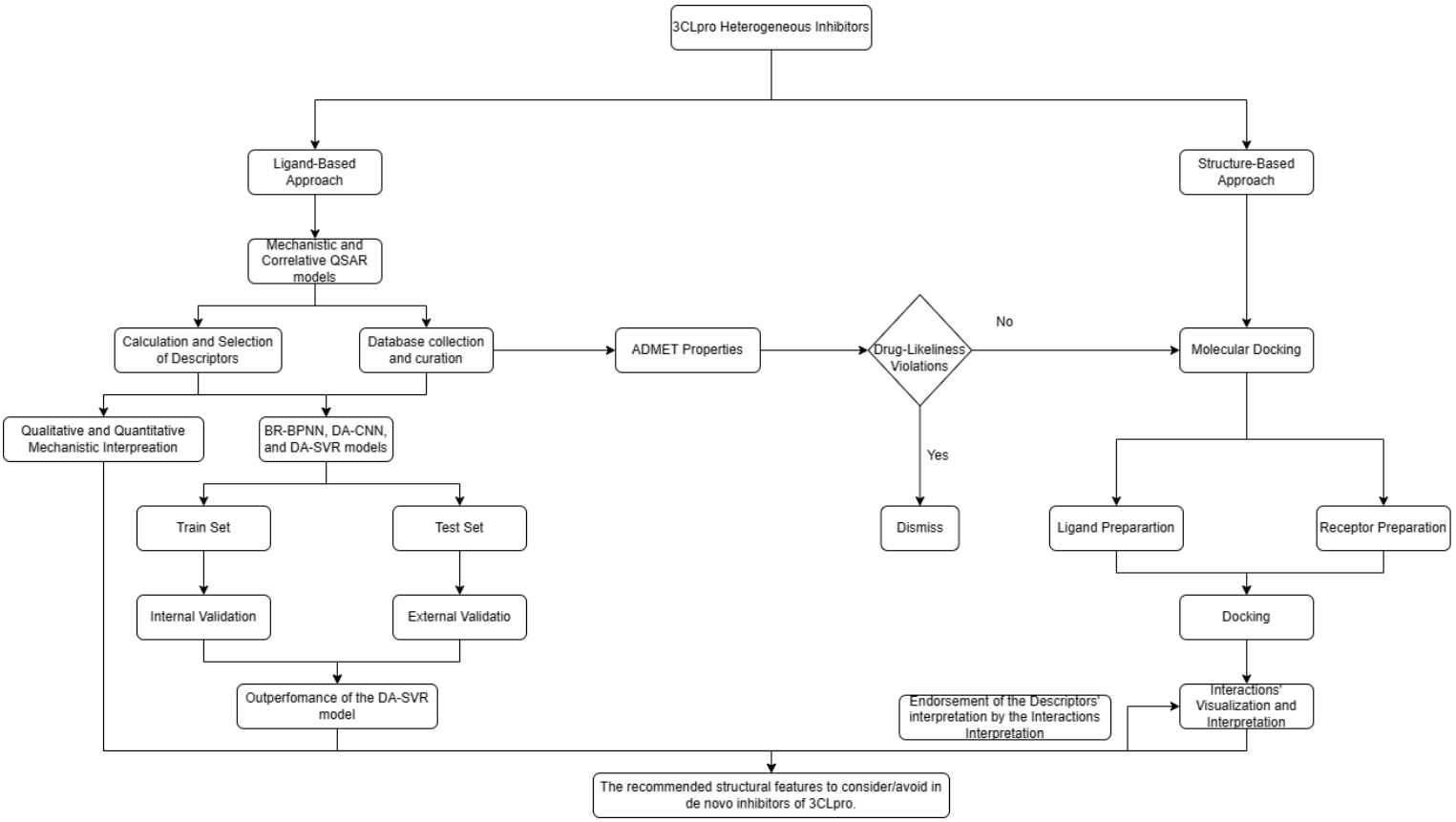
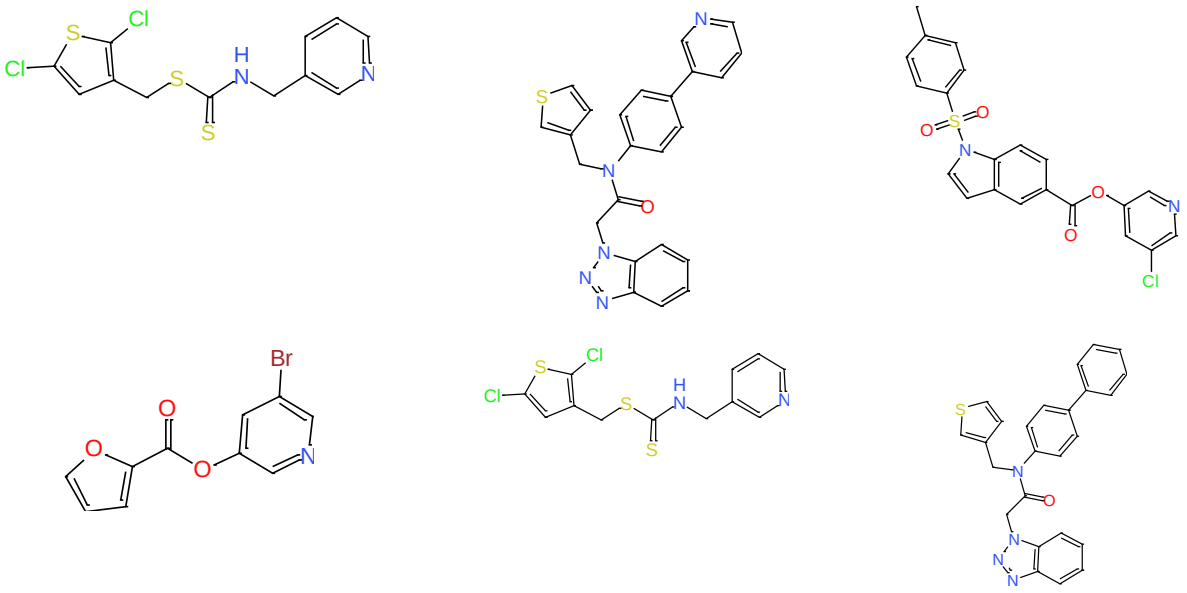


Fig III.3 organizational chart of the heterogeneous inhibitors modelling



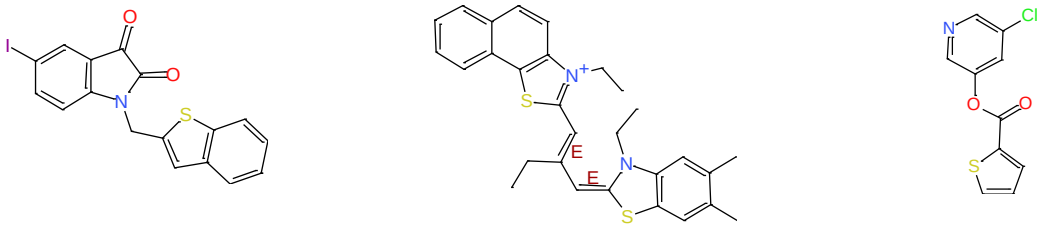


Fig III.4 Samples from the 246 selected inhibitors

2.2.4. Regression methods

2.2.4.1. Support vector machine

Details provided above. Please refer to 2.14.2. Support Vector Regression

2.2.4.2. Dragonfly Algorithm

First introduced in 2016, Dragonfly Algorithm DA is an intelligence optimization algorithm that is inspired by dragonflies' static and dynamic behaviors[352]. The optimization process relies on two major steps: exploration and exploitation achieved through attracting food sources and repelling enemies[203]. When coupled with the SVM regression method, DA allows the final hybrid model to be more performing, more robust, and less likely to overfit data thanks to its ability in solving nonlinear problems[400]. Hence, the DA algorithm was used to find the optimized combination of the hyper-parameters C, σ, and ε values for the SVR model.

2.2.4.3. Conventional neural network CNN

A Convolutional Neural Network (CNN) is a type of deep learning algorithm. CNNs in CADD are used to predicting biological activities and understanding molecular properties by automatically extracting intricate patterns from complex structures, thus significantly aiding drug discovery and design efforts. Its architecture involves the input, convolutional, pooling, fully connected, and output layers along with the learning rate and the activation function[401].

2.2.4.4. The hybridization of CNN and DA:

The Artificial Neural Network (ANN), specifically CNN, stands as a significant competitor to SVR in the realm of soft computing techniques[43]. Both of these established methods exhibit their own strengths and limitations within the domain of Computer-Aided Drug Design (CADD). Therefore, CNN was chosen for comparison with SVR, primarily due to its complexity and ability to handle intricate non-linear problems. To enhance CNN's performance, it was hybridized with DA to optimize the learning rate, a pivotal parameter in CNN's training and model compilation process. This coupling of CNN with DA serves two main purposes: firstly, exploring the uncharted territory of DA-CNN QSAR models in CADD, especially in predicting the activity of SARS-CoV-2 inhibitors, as no papers were published in this regard; and secondly, ensuring a statistically and algorithmically sound and comprehensive comparison between DA-SVR and CNN, necessitating the integration of CNN with DA.

2.2.4.5. Bayesian Regularization-Backpropagation Neural Network BR-BPNN

BR-BPNN is a type of Feedforward Neural Network FNN, a fundamental type of artificial neural network where information travels in only one direction –forward- from the input nodes, through hidden nodes (if any), and finally to the output nodes. Bayesian Regularization-Backpropagation" specifically refers to the training algorithm used for this neural network. In the case of BR-BPNN, Bayesian regularization is integrated with the backpropagation algorithm. Backpropagation is a supervised learning algorithm commonly used for training neural networks. It minimizes the error between the predicted output and the actual output by adjusting the network's weights iteratively[402]. The choice of the BR-BPNN was deliberate, leveraging its inherent power in a relatively simpler architecture. Unlike the complexity of SVR and CNN, BR-BPNN represents the fundamental and foundational form of Artificial Neural Networks ANN. Despite its basic structure, BR-BPNN proves to be exceptionally powerful[403]. Therefore, BR-BPNN was used to examine the level of complexity of the studied phenomenon. It was used solely without hybridization since it has only one parameter to tune, which is the number of neurons.

2.2.5. Validation metrics

Statistical validation of any QSAR model plays a key role in the reliability of the considered model[404]. This important step allows the assessment of the predictive ability of the model both internally (the training set) and externally (the test set)[211]. This is achieved through the calculation of several criteria listed in table Table III.1 namely the coefficient of determination R^2, coefficient of cross-validation Q^2, and the Root Mean Squared Error RMSE. The insights gained from these metrics allow the judgment of the performance and generalization of the model.

2.2.6. Applicability domain

Details provided above. Please refer to 2.1.6. Applicability domain

2.2.7. ADMET properties

Absorption, Distribution, Metabolism, Exertion, and Toxicity are the main reason why most drug candidates fail in clinical trials[269]. A significant barrier to drug discovery and design is the integrated optimization of the ADMET properties[268]. To predict the ADMET properties, the SwissADME online server[250] was used to determine (i) the drug likeliness with respect to the Lipinski Rule of 5, Veber, Ghose, Muegge, and Egan and (ii) lead molecules and synthesis accessibility scores. While the ADMET Lab 2.0 web server[271] was used to predict the Absorption, Distribution, Metabolism, Exertion, and Toxicity of the molecules. Furthermore, 24 molecules that showed good pharmacokinetics and drug-like properties with respect to the Lipinski rule, were kept for further docking analysis.

2.2.8. Molecular docking

2.2.8.1. Macromolecule preparation

The 3-dimensional crystal structure of the target protein (3CLpro) was retrieved from the trusted[270] Protein Data Bank PDB website of macromolecules' crystal structures[405]. The protein under code 6LU7 with a resolution of 2.16 Å was downloaded in PDB format. Undesirable molecules, such as water molecules, ligand along with extra chains were

removed and the protonation was performed using the UCSF Chimera package (version 1.16)[406]. The determination of the active site of the enzyme was performed using the previous software.

2.2.8.2. Small molecules preparation

The 24 molecules showing drug-likeliness properties were optimized by minimizing their energy using the MMFF94 forcefield method and generating their 3-dimensional conformers. The visualization, generation of the conformers and minimization of the energy were carried out using the Data Warrior software. The molecules were saved in a single file in SDF format and can be found in Table III.3.

2.2.8.3. Molecular docking analysis and post-docking protocol

Molecular docking stands for great importance in drug discovery as it studies the affinity between the target and the ligand at different poses of the ligand(Stanzione et al. 2021). The molecular docking was performed in Auto Dock Vina(Trott and Olson 2010) using the script-based multiple docking method in an Ubuntu environment.

Table III.3 Major ADMET properties for the selected 24 molecules

SMILES	**Lipinski Rule**	**Pfizer Rule**	**GSK Rule**	**Golden Triangle**	**Ghose Rule**	**Veber Rule**	**Muegge Rule**	**Leadlikeness Rule**
Clc1cncc(c1)OC(=O)C1=CC=C[C@H]2C1=NC=C2	Accepted	Rejected	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted
Clc1ccnc(c1)OC(=O)C1=CC=C[C@H]2C1=NC=C2	Accepted	Rejected	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted
NC(=O)c1ccc2c(c1)C(=O)C(=O)N2Cc1ccc2c(c1)cccc2	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted
Clc1cncc(c1)OC(=O)C1=CC2=CC=N[C@H]2C=C1	Accepted	Rejected	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted
O=C(C(=O)C1=CC2=CC=N[C@H]2C=C1)On1nnc2c1cccc2	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted

Nc1ccc(cc1)c1nc2c(s1)c(c(cc2)C)S(=O)(=O)O	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted
Oc1cc(O)c2c(c1)oc(c(c2=O)O)c1ccc(c(c1)O)O	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted
CC1CCN(CC1)S(=O)(=O)c1ccc2c(c1)C(=O)C(=O)N2C	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted
COc1ccc(cc1O)[C@@H]1CC(=O)c2c(O1)cc(cc2O)O	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted
CC1CC(C)CN(C1)S(=O)(=O)c1ccc2c(c1)C(=O)C(=O)N2	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted
N#Cc1ccc2c(c1)C(=O)C(=O)[C@H]2Cc1c(C)noc1C	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted
CC1CC(C)CN(C1)S(=O)(=O)c1ccc2c(c1)C(=O)C(=O)N2C	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted
OC(=O)C(=O)CC(=O)c1cccc(c1)NCc1ccccc1	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted
O=C1Nc2c(C1=O)cc(cc2)S(=O)(=O)N1CCCCC1	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted
C[C@H]1CCCCN1S(=O)(=O)c1ccc2c(c1)C(=O)C(=O)N2	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted

FC1=C[C@H]2C(=CC(=N2)C(=O)C(=O)On2nnc3c2cccc3)C=C1	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted
Clc1ccc(cc1)S(=O)(=O)c1cnc(nc1N)N	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted
O=C(CN(S(=O)(=O)c1cccc(c1)Cl)C)NC1CCCC1	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted
Brc1cncc(c1)OC(=O)c1ccco1	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted
Clc1cncc(c1)OC(=O)C1=CC=CC2=NC=C[C@H]12	Accepted	Rejected	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted
Clc1ccnc(c1)OC(=O)C1=CC2=NC=C[C@H]2C=C1	Accepted	Rejected	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted
Clc1cncc(c1)OC(=O)C1=CC2=NC=C[C@H]2C=C1	Accepted	Rejected	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted
CC1CCN(CC1)S(=O)(=O)c1ccc2c(c1)C(=O)C(=O)N2	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted
O=C1N(C)c2c(C1=O)cc(cc2)S(=O)(=O)N1CCOCC1	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted

The visualization of the docking results in terms of the best ligand pose and the examination of the target-ligand interactions was performed by the BIOVIA Discovery Studio 2021 tool. It illustrates the 2 and 3-dimensional interactions with the bond's lengths. Ligands were examined separately by BIOVIA DS in a PDB file containing the enzyme complexed with the ligand in the best pose.

CHAPTER IV
Results and Discussion

1. ML300 analogues

1.1. Selection of representative descriptors

The selection was performed using the GA in Build-QSAR software, revealing the seven most representative descriptors: numberOfHal, SPC-6, WTPT2, MOMIXZ, KierShape2, Numberrotatablebonds, ATSm2.

1.1.1. Constitutional descriptors

Number of hal: A 1-dimensional descriptor representing the number of halogen atoms in the molecule. Halogens are highly electronegative species and are thus highly reactive as they interact with a donor of electron density[407]. The organic halogens form halogen bindings with the amino acid residues in the active site, specifically chlorine that highly interacts with leucine[408].

Number rotatable bonds: It is a one-dimensional descriptor that counts the number of bonds that allow free rotations around themselves[409]. The existence of rotatable bonds increases the bioavailability of the molecule and thus indicates the ability for good drug efficacy. The bioactivity depends on the number of rotatable bonds: ten rotatable bonds or fewer bonds have a significant positive impact on the bioavailability and, henceforth, on the molecule's activity[410].

1.1.2. Topological descriptors

Topological descriptors are determined from the graphical representation of molecules without hydrogen atoms. They are thus two-dimensional descriptors that value the weighted arrangement of atoms and their bonds[322].

SPC-6: In general, connectivity indexes measure the possibilities of a molecule interacting with another molecule in a milieu. SPC-6 is a 2-dimensional descriptor that stands for Chi Path/Cluster index. It evaluates the connectivity index at the sixth order or the sixth degree of branching[325]. SPC-6 considers an atom's number of electrons and valence electrons. It represents the hypervolume of order 6 in a molecule resulting from contributions of cluster/path fragments[411] and, therefore, the accessible reactivity space of a molecule with the target, in this case, the active site of 3CLpro.

WTPT2: Weighted path-2 is a two-dimensional descriptor[412] that measures weighted paths of length two in a molecule divided by the number of atoms in the same molecule[412]. In graph theory, a path is an alternating sequence of vertices and edges that begins and ends with a vertex without passing by the same vertex more than once. A weighted path assigns a numerical value (weight) to each edge depending on the type of atom existing, thus providing information about the alignment and sequence of atoms and, henceforth, about the physicochemical properties of a molecule[412]. Additionally, molecules with similar WTPT2 values are more likely to share similar structures and, thus, similar potency. Therefore, WTPT2 can be very useful in similarity screening. More importantly, high values of WTPT2 indicate a more compact and branched structure, with shorter average distances between pairs of atoms. This is useful in inhibiting 3CLpro since the latter has multiple sub-sites and a

complex shape with several pockets. A more branched complex molecule is likelier to fit and interact with the 3CLpro active site[413].

Kier Shape 2: It calculates the Kier shape for paths with a length of two[414]. This descriptor provides information about the molecular shape, normalized by the number of atoms relative to the star and linear graphs: the extreme shapes a molecule can take. The shape of a molecule directly influences the structural complementarity with the active site of the target enzyme. It allows the occurrence of intermolecular forces with minimum energy. K2 index quantitates the contribution of the molecular shape to the molecular activity.

1.1.3. Geometrical descriptors

MOMIXZ: A three-dimensional descriptor that measures the moment of inertia along the axes ratio X to Z[414]. The moment of inertia generally describes how the mass of a rigid body is distributed. The descriptors MOMY X and MOMY Z measure how hard it is to spin the molecule along the axes X and Z, respectively. The closer the mass is to the axis, the smaller the moment of inertia. MOMIXZ compares the moment of inertia of axis X to Z. In other words, it provides information about the molecule's orientation when in contact with the active site of 3CLpro. It is also noted that the moment of inertia evaluates properties such as geometry, shape, and conformation[415]. This stands for great importance as the activity of inhibitors is directly related to the spatial conformation complementarity between the enzyme's active site and the molecules. A molecule must be in the proper position and have the corresponding conformation to exhibit its inhibitory effect correctly. Otherwise, it cannot attach to the enzyme.

The analysis of the descriptors indicates that the nature of atoms and their sequence, bonds, and shapes influence their activity. Considering the properties of the active site of 3CLpro, it is logical that the nature of atoms, represented by the number of halogens, matters because it is the engine of all kinds of bonds between the enzyme and the inhibitors. The molecules' planetary (rotatable bonds, connectivity, shape, autocorrelation, weighted path) and 3-D dimension (moment of inertia) reflect the aspect of structural complementarity, as they consider the shape, the sequence, the neighbouring skeletons, and the rotation among axes.

The contributions of each descriptor are illustrated in Fig IV.1 The relative importance of the selected descriptors was calculated using the Mean Effect method[416]. The MOMIXZ descriptor had the greatest positive importance, explaining 40% of the activity, followed by the Kier Shape 2, Numberrota table bonds, ATSM2, and finally WTPT2, the SPC-6 had the most significant negative contribution, followed by the Number of hal, with the most negligible relative importance. The results further sustain the above interpretation, indicating that to maximize the inhibitory activity of these nonpeptidic noncovalent molecules, the structural complementarity in terms of mass centered towards the X axis over the Z axis, the presence of rotatable bonds at least 20% in the molecules, the weighted paths of lag 2 and the shape of fragments of lag 2 influence positively the activity. Whereas the number of halogens slightly decreases the activity, the significant branching of molecules (6th degree) hinders the binding with the active site. This explains the inversely proportional relationship governing the degree of complexity of the hypervolume of interaction and the activity.

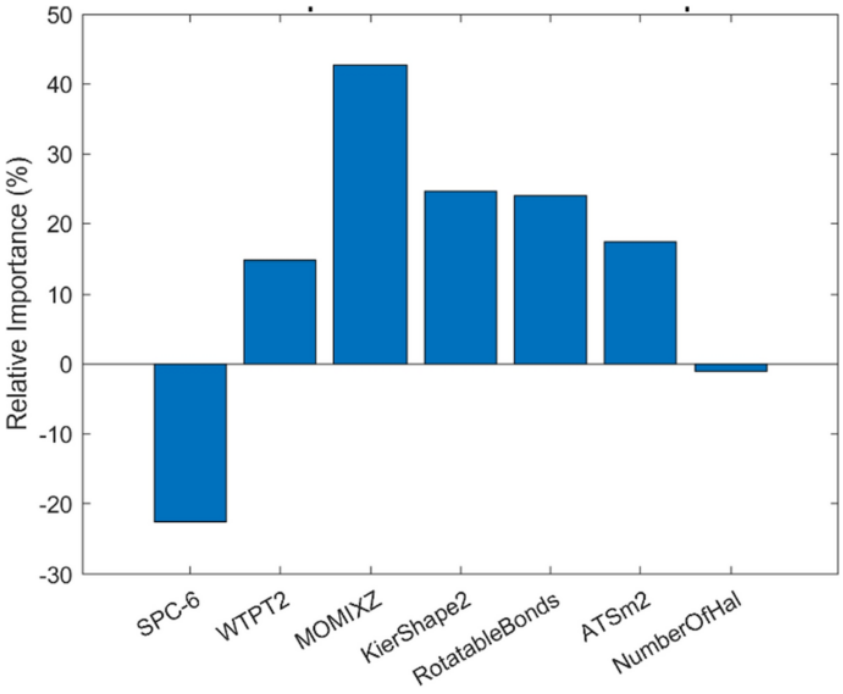


Fig IV.1 The relative importance of the selected descriptors

1.2. QSAR models' development

1.2.1. MLR model

This study developed an MLR model to predict the inhibitory activity of various molecules against SARS-CoV-2, the 3CLpro viral enzyme. The established MLR model (Eq. 1) shows the linear relationship between the selected descriptors and the activity.

$$\begin{aligned} \mathrm{pIC50} = {} & 4{,}11278 + 0{,}17372\,\mathrm{SPC6} - 0{,}50349\,\mathrm{WTPT2} - 1{,}27647\mathrm{MOMI\,XZ} - \\ & 0{,}15793\mathrm{types,KierShape2} - 8{,}14971\mathrm{types,FractionRotatableBonds} - \\ & 0{,}02892\mathrm{ATSm2} \end{aligned} \qquad \text{Eq. 1}$$

The coefficients linked to each descriptor represent the weights and the type of contribution (positive or negative). The validation criteria in Table III.2 for this model showed its weak ability to predict with relatively low R and R^2 values (under the accepted value of $R^2 < 0.6$). The table indicates that the MLR model failed to pass the threshold value for $\bar{r}m^2$ (> 0.5) and Δrm^2 (< 0.2)[335]. Fig IV.2 shows the regression plot for the train and test sets, where the predicted values do not corroborate with the experimental values. The highest AIC value of the model shown in Table IV.1 (-64.9684) indicates that it fitted the experimental data very weakly. The experimental and predicted values by the MLR model are presented in Table S-IV.1 in the annex.

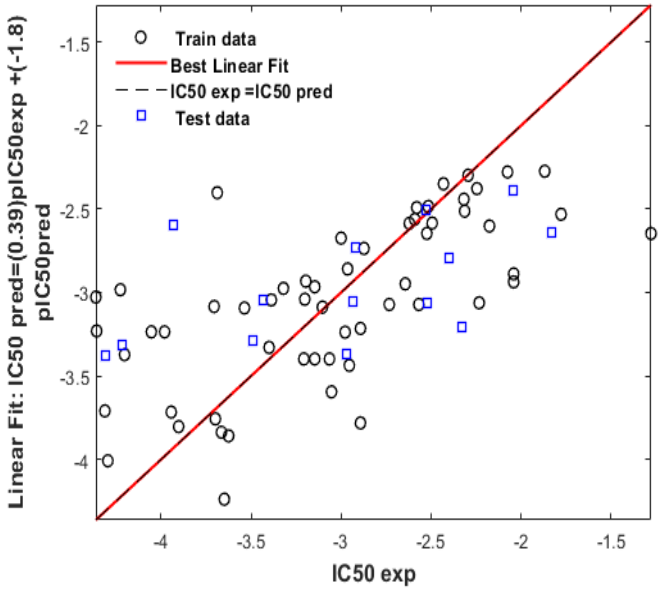


Fig VI.2 MLR regression plot

Table IV.1 Statistical criteria for the MLR, SVR, and PSO-SVR models

Validation criteria	**Model**		
	MLR	**SVR**	**PSO-SVR**
$\mathbf{R^2}$	0.4350	0.9665	0.9720
$\mathbf{Q^2}$	0.4350	0.9664	0.9720
$\mathbf{\bar{r}_m^2}$	0.2723	0.9497	0.9550
$\mathbf{\Delta r_m^2}$	0.3254	0.0283	0.0244
RMSE	0.5540	0.1367	0.1243
AIC	-64.9684	-249.1234	-264.5433

1.2.2. Support vector regression model

To determine the optimal values of the SVR hyperparameters, a loop was constructed to evaluate the cost function values each time based on the hyperparameters' incrimination. The chosen Kernel function and the best combination of the hyperparameters were found to be $C = 500$, $\gamma = 1$, $\varepsilon = 0.0485$, and the Radial Basis Function RBF as the kernel function. The predicted values of pIC_{50} by the SVR method are listed in Table S-IV.2 in the annex. The evaluation criteria for the training set and the test set performance of the SVR model are summarized in Table 2. The predicted values of pIC_{50} of the inhibitors for the training set test set versus their observed values are represented in Fig. IV.3. The developed model showed a good correlation between the predicted and observed values with low RMSE (0.1367), high

R2 (0.9665), and good robustness indicating by the Q2, $\bar{r}m2$, and $\Delta rm2$ values that satisfy the threshold value. The AIC value of this model (−249.1234) reveals that the SVR model fitted the experimental data better than the MLR model.

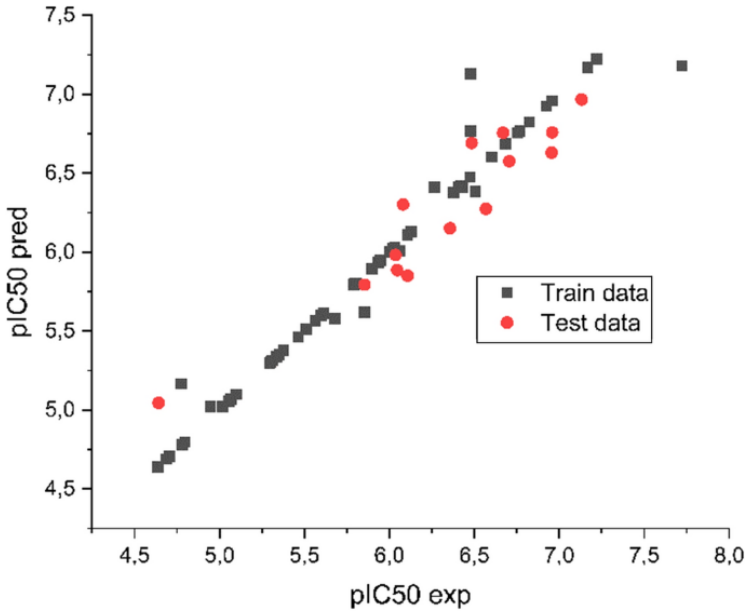


Fig IV.3 SVR Regression plot

1.2.3. SVR-particle swarm optimization model

The combination of the PSO and SVR algorithms aims to find the optimal variety of the hyperparameters that yields the smallest error values by testing a huge number of possible combinations. The best combination consisted of C = 131.7001, γ = 1, and ε = 0.0350. It is thus a way to avoid the overfitting of the model with a higher R2 coefficient (0.9720) and lower RMSE value (0.1243). The values of Q2, $\bar{r}m2$ and $\Delta rm2$ of this model also satisfy the threshold value. The value of R2-Q2 for the whole set (both training and test) 0.9650–0.9649 is 0.0001 (accepted value < 0.3), indicating that the model is not overfitted[206]. Table IV.1 shows the statistical parameters for the PSO-SVR model. The AIC value of this model (−264.5433) indicates that the model could fit the experimental data very well. Fig IV.4 illustrates the correlation between the predicted pIC_{50} and the observed values for the training and test sets.

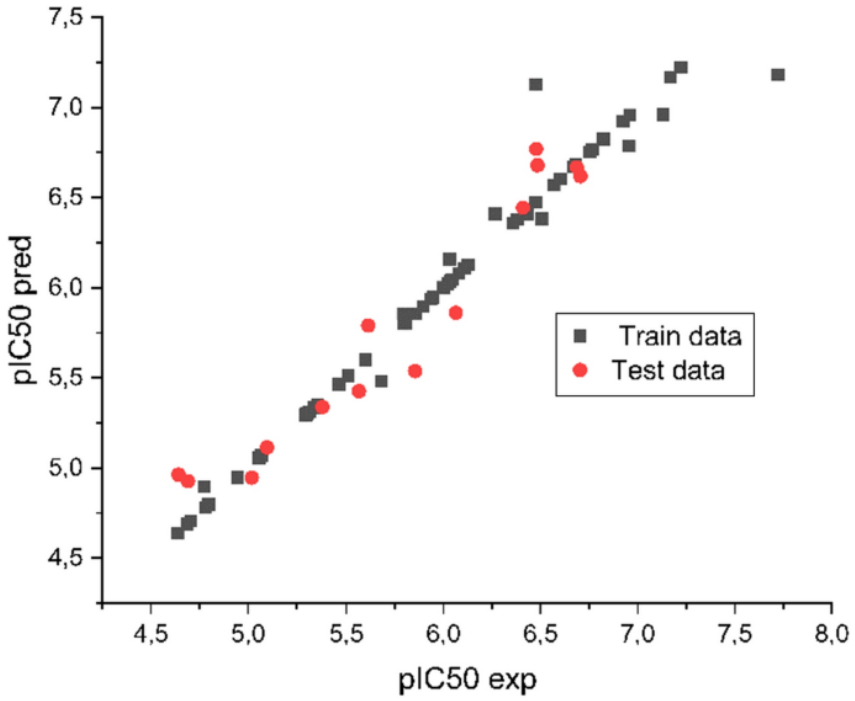


Fig IV.4 PSO-SVR Regression plot

Table IV.2 The experimental and predicted values of pIC_{50} predicted by the PSO-SVR model, train values in pink, and test values in green.

pIC50 experimental values	Predicted by SVR-PSO
-4,361727836	-4,361784123
-4,35755372	-4,075910866
-4,311753861	-4,311766288
-4,309417226	-4,309393707
-4,295127085	-4,295536343
-4,225825991	-4,225552995
-4,202488317	-4,202273317
-4,053462605	-4,053522706
-3,980003372	-4,053522706
-3,930949031	-3,931069181
-3,902002891	-3,902231422
-3,705007959	-3,705040286
-3,687528961	-3,687458966
-3,664641976	-3,664655325
-3,648360011	-3,648372176
-3,624282096	-3,624294261
-3,489958479	-3,489583069
-3,435047641	-3,435038271
-3,399673721	-3,399909143
-3,387389826	-3,387262358
-3,320146286	-3,560089118

-3,201397124	-3,201044374
-3,195899652	-3,19578758
-3,146128036	-3,146175938
-3,103803721	-3,103769464
-3,053078443	-3,053181556
-3	-2,999528784
-2,977723605	-2,978069675
-2,968482949	-2,858332487
-2,963787827	-2,96388628
-2,954242509	-2,954097154
-2,936513742	-2,936782252
-2,919078092	-2,918879698
-2,892094603	-2,892107029
-2,892094603	-2,891934386
-2,872738827	-2,872392347
-2,73239376	-2,723830484
-2,621176282	-2,621561383
-2,591064607	-2,591244671
-2,579783597	-2,580219499
-2,568201724	-2,723791496
-2,523746467	-1,907639805
-2,522444234	-2,522099604
-2,494154594	-2,624791863
-2,431363764	-2,431474917
-2,397940009	-2,398079074
-2,330413773	-2,330132903
-2,318063335	-2,31762089
-2,245512668	-2,245555943
-2,176091259	-2,175690512
-2,075546961	-1,92299062
-2,041392685	-2,041530741
-2,041392685	-2,041356723
-1,86923172	-1,869318468
-1,832508913	-1,832608239
-1,77815125	-1,778012686
-1,278753601	-1,821768458
-2,045322979	-2,42889648
-2,31386722	-2,385184222
-2,526339277	-2,655330969
-4,218797998	-4,164205024
-2,23299611	-2,521138734
-2,294466226	-2,295267553
-2,515873844	-2,349846077

-2,643452676	-2,802084604
-3,064457989	-3,09915499
-3,146128036	-3,496282563
-3,206825876	-3,146175938
-3,536810866	-3,63050825
-3,698100546	-3,668349001
-3,942999593	-3,864053299

1.2.4. External validation

The calculation of the external validation criteria on the test set of each model shows the performance for SVR and PSO-SVR followed by MLR (R^2: 0.9310, 0.8932, and 0.2851respectively) and low RMSEP values (0.1870, 0.2203, and 0.6465 respectively. The external validation criteria evaluated the models' accuracy and ability to predict unseen data, and the results are demonstrated in Table IV.3. The results show that the MLR model could not perform well on the test data. This is shown by the low CCC value, the high root mean square error RMSE and the mean absolute error MAE. The PSO-SVR model outperformed the SVR model regarding generalization and prediction ability regarding correlation metrics and error metrics, indicated by higher CCC and Q_2F_1 values and lower RMSE and MAE values. Generally, the SVR and PSO-SVR models showed good statistical significance and robustness.

Table IV.3 External validation criteria for the MLR, SVR, and PSO-SVR models

Validation criteria	**Model**		
	MLR	**SVR**	**PSO-SVR**
$\mathbf{R^2_{Test}}$	0.2851	0.8932	0.9310
$\mathbf{Q^2_{F1}}$	0.6327	0.9121	0.9765
$\mathbf{\bar{r}^2_m}$	0.0414	0.7671	0.8959
$\mathbf{\Delta r^2_m}$	0.4245	0.1099	0.0569
CCC	0.3877	0.9218	0.9639
RMSE	0.6465	0.2203	0.1870

1.2.5. Applicability domain

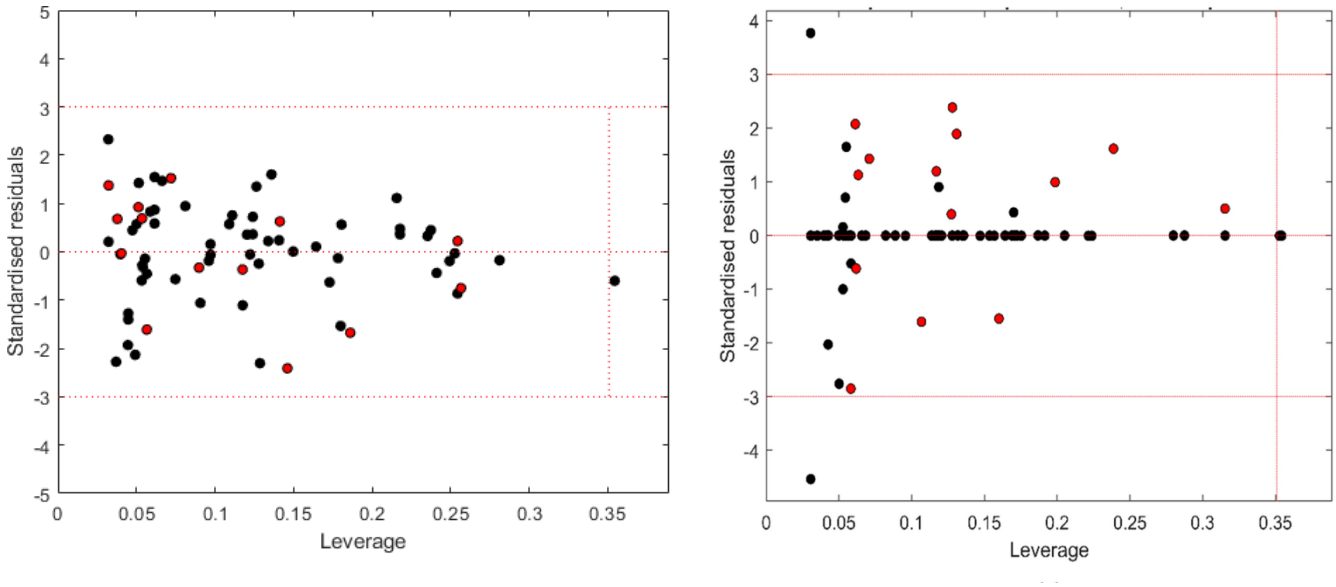
Multiple linear regression

The domain of applicability of the MLR model was analyzed by the William plot (Fig IV.5.a). It shows a total distribution of the predicted points in the range of the.applicability domain

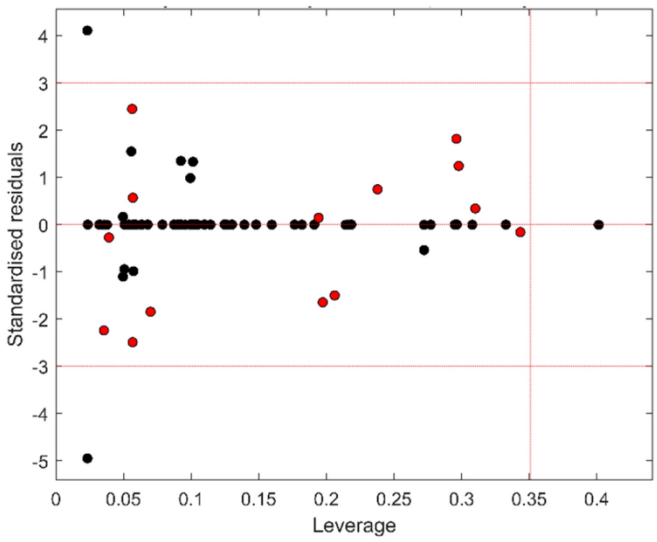
[−3, 3], with one influential point exceeding the critical leverage point of 0.35. All the predicted values did not align close, but all fell inside the domain, indicating that the model found no response outliers

Support vector regression

The applicability domain was analyzed using the William Plot, as shown in Fig IV.5.b. This shows that about 98.59% of the values are within the applicability domain's critical leverage point range. Also, it was noticed that the whole test set and 96.5% of the training set are included in the domain. Only two points (molecules 1 and 14) are outside the domain with relatively high standardized residual values.



(a) (b)



(c)

Fig IV.5 Applicability domain of (a) MLR model, (b) SVR model, and (c) PSO-SVR model

SVR-particle swarm optimization

William's plot for the PSO-SVR model (Fig III.7.c) shows that one molecule (molecule 55) did exceed the critical leverage value of 0.35, and two other training molecules (molecules 1 and 17) lie outside the standard residual's boundaries [−3, 3]. However, most points aligned with the 0 standard residuals, and only a few were distributed in the space domain.

1.2.6. Models' comparison

The analysis of the applicability domains indicates that despite the existence of a few outliers in the SVR and PSO-SVR models, they remain reliable as the standard residual values of the influential molecules are low, and the majority of the points align with the zero standards residual rather than covering the whole domain. A closer look at the SVR and PSO-SVR models shows that the predicted values by the PSO-SVR align better with zero standard residual value than those predicted by the SVR, and both models had an X outlier. These observations suggest that molecules 55 and 34 for PSO-SVR and SVR models are descriptors' outliers and, thus, an influential point. Although the presence of outliers is a normal phenomenon, it may be due to the model errors and/or errors associated with experiments. In conclusion, the analysis of the applicability domain confirmed the robustness and the predictivity of the PSO-SVR model.

According to the statistical criteria above, the PSO-SVR model showed the best performance, followed closely by the SVR model with good predictivity. In contrast, the MLR model had relatively weak metrics, resulting in the lowest performance and predictability. While the MLR method aims to build a predictive model, it is also, more importantly, to verify the linearity of the studied phenomenon. It confirms that the studied phenomenon is nonlinear and that the MLR method cannot be applied to this database. Henceforth, using nonlinear methods is a justified need, which explains the better results obtained by the PSO-SVR and SVR algorithms. Using a metaheuristic algorithm to tune the SVR parameters rather than the trial-and-error method is one reason the PSO-SVR method performed better. Moreover, it avoids the overfitting of the model as it allows the model to reach a degree of complexity suitable to that of the studied phenomenon.

Moreover, the SVR model showed a close performance on the training set with the PSO-SVR model with $R^2SVR = 0.96$, $R^2PSO\text{-}SVR = 0.97$, $RMSESVR = 0.12$, and $RMSEPSO\text{-}SVR = 0.13$. This renders it complex to conclude the outperformance of PSO-SVR. However, a closer look at the external validation criteria shows the significant outperformance of PSO-SVR with $R^2 = 0.93$, $Q^2 = 0.97$, and $RMSE = 0.18$ in contrast to $R^2 = 0.89$, $Q2 = 0.91$ and $RMSE = 0.22$ of the SVR model, indicating the superiority of PSO-SVR in generalization and predictive ability of unknown datasets, endorsing the importance of parameters tuning by metaheuristic algorithms. These findings are confirmed by the analysis of the AIC, where the PSO-SVR model had the lowest value, followed by SVR and MLR, indicating that the PSO-SVR performed the best. Furthermore, since the difference in the AIC values between the PSO-SVR and SVR models is greater than 2, the PSO-SVR model can safely be considered significantly better than the SVR model, indicating that having a metaheuristic algorithm improved the prediction results.

1.2.7. Comparison with literature

Many papers were published to provide a predictive model for the 3CLpro inhibitors, using classical algorithms, with acceptable performance. The detailed comparison is demonstrated in Table IV.4. However, the hybrid PSO-SVR algorithm improved the models' precision and accuracy and outperformed all the listed models. This confirms the significance of the proposed model in predicting the inhibiting activity against the 3CLpro.This outperformance can be algorithmically explained as follows:

Nonlinearity aspect: Unlike MLR, PLS and their hybridization (GA-MLR and PLS-MLR), which apply only to linear problems, SVR can handle complex nonlinear issues with excellent efficiency. Notably, the prediction of the inhibitory activity of molecules is a complex phenomenon, most often nonlinear, and thus requires a strong nonlinearity potential and the ability to adapt to its level of complexity, which linear algorithms fail to have. This is why SVR outperformed linear models such as MLR and PLS.

Generalization and overfitting: SVR and ANN have their strengths and weaknesses; SVR can be considered better than ANN when applied to small datasets. ANN tends to overfit when trained on a small dataset since it has many parameters. When the number of parameters is large enough compared to the number of observations, ANN tends to learn the noise instead of the underlying patterns, which affects its performance on the test/external datasets. On the contrary, SVR has fewer parameters to tune and regularization parameters, ensuring avoiding overfitting problems and a good performance on unseen datasets[417]. In Computer-Aided Drug Design (CADD) and QSAR modelling for identifying SARS-CoV-2 inhibitors (or a novel pandemic or disease), little is known about potential molecules, and derivatives and analogues are designed in a limited number. Therefore, datasets of molecules against novel diseases like COVID-19 tend to be small and risk being overfitted if modelled by ANN. This explains why the proposed model outperformed the ANN-reported model.

Flexibility and ability to handle outliers: Hologram QSAR (HQSAR), Comparative Molecular Field Analysis (CoMFA) 3D QSAR and Monte Carlo techniques are based on different principles (images of electronic structures and steric properties of molecules for HQSAR[371], deriving quantitative relationships between the activity and the structure by comparing the electrostatic and steric interactions of molecules and a reference drug for CoMFA[372], and evolving randomly to generate several models and selecting the best one for Monte Carlo[369]). However, they are similar in how they differ from SVR: they use images and interactions as input instead of comparison or simulation instead of numerical values of molecular descriptors and rely on comparing molecular similarity or fitting a mathematical function to data instead of supervised machine learning. SVR uses less complex inputs and can adapt to different levels of complexity thanks to the kernel functions and perform better on nonlinear phenomena. Also, thanks to the support vectors and the boundaries in SVR, it handles outliers more effectively than the methods above, which can be very sensitive to outliers. This is useful, especially when handling heterogeneous datasets.

PSO robustness: Particle Swarm Optimization (PSO) is used to tune the parameters of machine learning algorithms. It is advantageous over other metaheuristic algorithms such as

Genetic Algorithm GA since it is less sensitive to the initialization parameters, unlike GA, which is highly influenced by them and, therefore, less robust than PSO. PSO allows a large search space leading to a fine-tuning of parameters and, thus, a better convergence, in addition to a short computational time. Algorithmically, this is the reason why the proposed PSO-SRV model outperformed the reported GA-SVR model[418].

Table IV.4 Comparison between literature and the present paper

References	**Target**	**Model output**	**Best model**	**Database size**	**Metrics**
[364]	3CLpro	Ki	SVM-GA	89	R^2= 0.814 RMSE= 0.459
[365]	3CLpro	BFE	MLR	226	R^2=0.80 RMSE=0.52
[366]	3CLpro	pIC_{50}	MLR	73	R^2=0.73 Q^2=0.86
[120]	3CLpro	pIC_{50}	PLs-MLR	104	R^2=0.75 Q^2=0.70
[367]	3CLpro	pIC_{50}	MLR	41	R^2=0.69 RMSE=0.32
[368]	3CLpro	pIC_{50}	ANN	90	R^2=0.86 RMSE=0.33
[369]	3CLpro	IC_{50}	Monte Carlo	40	R^2=0.91 Q^2=0.68
[419]	3CLpro	EC_{50}	GA-MLR	37	R^2=0.84 RMSE=0.21
[370]	3CLpro	pIC_{50}	KPLS	33	R^2=0.81 RMSE=0.56
[371]	3CLpro	pIC_{50}	HQSAR	69	R^2=0.86 Q^2=067
[372]	3CLpro	pIC_{50}	CoMFA	23	R^2_{train}=0.75 Q^2=0.64

[267]	3CLpro	$pMIC_{50}$	GA-MLR	25	R^2_{train}=0.75 RMSE=0.14
This study	3CLpro	pIC_{50}	PSO-SVR	71	R^2=0.96 R^2_{train}=0.93 Q^2=0.96

While this outperformance has algorithmic reasons, it is worth clarifying that comparing different algorithms is very complex and requires a thorough, comprehensive analysis of various factors, mainly database similarity, size and quality. The comparison was elaborated for indicative purposes and was based on the similarity of the target 3CLpro solely to position the proposed SVR-PSO in the range of errors. Therefore, it does not imply the absolute superiority of SVR-PSO over other algorithms but rather in this proposed context.

2. Heterogeneous datasets:

2.1. Descriptors' mechanistic interpretation

The selection of the pertinent descriptors revealed a set of 10 representative descriptors shown in Table IV.5. The mechanistic interpretation of the selected descriptors is a hard task,. especially for some pure mathematical descriptors such as eigenvalues, spectral and E-state descriptors[420] Moreover, the extraction of the Structure-Activity Relationship SAR is not a direct process when it comes to heterogeneous datasets (like the dataset considered in the present study), unlike homogenous datasets, which share the same pharmacophore and mechanism of action[420]. Hence, we consider the extraction and evaluation, both in quantity and in quality, of the SARs encoded in the selected descriptors.

A spectral diameter is a difference between the largest and the smallest eigenvalues[421]. The largest eigenvalue is proportional to the mean of the number of neighboring atoms to atom i; while the smallest eigenvalue corresponds to the Highest Occupied Molecular Orbital HOMO which is directly involved in molecular ionization, resulting in the modification of the reactivity of the molecule. Thus, the more negative the HOMO eigenvalue, the more reactive a molecule is. It also decreases with high branching degrees and the presence of side chains[422]. Thus, for an increase in the spectral diameter, a high degree of branching (neighboring atoms) and a low HOMO eigenvalue are needed to make an inhibitor more reactive. It can be deduced that the relation between the spectral diameter and the powerfulness of an inhibitor is proportional.

An augmented edge adjacency matrix is a weighted adjacency matrix (nonzero diagonal adjacency matrix) that distinguishes edge sets (bond classes) from each other[423]. SpMAD(dm) is weighted by dipole moment, and thus characterizes bonds by their dipole moment.SpMAD is the absolute difference between an edge I eigenvalue and the average

eigenvalue (the absolute deviation) divided by the number of edges (and thus the word mean). This quantity is size-independent[421]. A bond dipole moment is the measure of the bond polarity between two atoms. Hence, weighting this spectral property by the dipole moment reveals its distribution in the edge sets reflected by the eigenvalues, regardless of the size of vertices forming the edge. The greater the SpMAD_AEA(dm), the greater the eigenvalue of the considered bond is over the average eigenvalue, the greater the dipolar moment of that bond, the more electronegative it is, and thus the more reactive the bond is. Therefore, it is a measure of the reactivity of different sets of bonds within the inhibitor.

Eig02_AEA(dm) is the second eigenvalue from the augmented edge adjacency matrix weighted by the dipole moment. In algebra, the second eigenvalue often risks being too close to the first largest eigenvalue as there is no proper upper bound for it. The Eig02_AEA(dm) should then represent the distribution of the dipole moment along the bonds present in an inhibitor. It is quite ambiguous and subtle to directly link this descriptor to a structural meaning, but mathematically speaking, considering the second largest eigenvalue will overshadow other eigenvalues and cause the leading eigenvalue to be repeated[424]. Therefore, an inversely proportional relationship between this descriptor and the SpMAD_AEA(dm) descriptor should exist, as the repetition of the leading eigenvalue with cause the average eigenvalue to increase, thus the difference between the single bond eigenvalue the average eigenvalue will diminish, hence diminishing the value of the SpMAD_AEA(dm).

nCbH is the number of unsubstituted benzenes in an inhibitor. Unsubstituted benzenes are well known for their high degree of unsaturation, and thus are very reactive species and are highly reactive with electron-deficient species. The histidine residue in the dimer of the 3CLpro active site acts like a base[425], and thus can potentially react highly with the unsubstituted benzenes in inhibitors. The descriptor nCbH representing the number of unsubstituted benzene rings is also present in molecules 19 and 15 (each containing 2 rings) and is also the motor of pi-bonds.

nCconj descriptor reflects the number of nonaromatic carbon atoms that have conjugated bonds. A bond is conjugated when there is an alternate of double and single bonds and concerns carbon atoms that have two single bonds and one double bond. According to Keeler et al.[426], the conjugation of bonds causes the delocalization of a lone pair of electrons decreases the reactivity. This means that the more an inhibitor has nonaromatic conjugated carbon, the less likely it is to react with the 3CLpro enzyme.

An alpha carbon is the carbon atom to which is attached a functional group. A hydrogen atom attached to an alpha carbon is called alpha hydrogen, referred to as the descriptor H-051. The presence of alpha hydrogens induces the phenomenon of conjugation to occur since the pair of electrons forming the bond between the alpha carbon and the alpha hydrogen are transformed back and forth to the carbocation. And thus, the more alphas hydrogens, the less reactive the alpha carbon. This agrees with the previous descriptor (nCconj) since nonaromatic conjugation decreases the potential of inhibitors.

The three electrotopological state descriptors encode the electronic and the topological environment of carbon atoms forming three aromatic single bonds, nitrogen atoms forming three single bonds, as well as alcohols[323]. Moreover, these descriptors show the most influential parts of the inhibitor on its activity, as they reflect the electronic richness (electronegativity) of a specific atom influenced by its adjacent atoms and the types of bonds the considered atom formed[323]. Thus, it is revealed that the electronic richness of Nitrogen atoms with three single bonds is important to the inhibition of the 3CLpro enzyme since such nitrogen is found in tetrapeptide and tetra peptidomimetic inhibitors (very famous in inhibiting 3CLpro), allowing them to perform nucleophilic attacks on the amino acids residues in the active site of 3CLpro[427]. The influence of the electronic state of carbon atoms with three aromatic single bonds (which are more likely to belong to benzenes) makes perfect sense as it was demonstrated that the number of unsubstituted benzene carbons was of positive impact on the inhibitory potential of the selected inhibitors. While nCbH previously interpreted takes account of the number of benzenes, minaaaC considers the electronic richness of these carbon atoms. On the other hand, the electronic state of terminal and/or radical alcohols reflects a potential reactivity between the alcohol functional group with the leucine present in the active site of 3CLpro[428].

A molecule is chiral when it is not superposable on its mirror image, and a chiral center is an atom with four or more different substituents[429]. Normally, chirality is a feature that requires 3D structures. however, advances in the QSAR and molecular descriptors could encode such a 3D property only from 2D structures[429]. The descriptor nLevel7 is an environment-related chirality descriptor that considers the number of atoms that are 7 bonds far away from the chiral center (a radius of 7). nlevel descriptors were found to be good indicators of the presence the chiral cliffs, which affect the recognition process and the binding affinity of a ligand towards the target, and thus help identify the best molecule in a chiral pair[429]. This can also be an indication of the presence of multiple side chains, which can potentially interact with each other rather than with the enzyme target, thus decreasing the interactions of such complex inhibitors with the target.

Table IV.5 The selected molecular descriptors

Descriptor	**Class**	**Description**	**Reference**
SpDiam_A	2D matrix-based descriptors	spectral diameter from the adjacency matrix	[324]
SpMAD_AEA(dm)	Edge adjacency indices	spectral mean absolute deviation from the augmented edge adjacency matrix weighted by the dipole moment	[324]
Eig02_AEA(dm)	Edge adjacency indices	eigenvalue n. 2 from augmented edge adjacency matrix weighted by the dipole moment	[324]
nCbH	Functional group counts	number of unsubstituted benzene C(sp2)	[324]

nCconj	Functional group counts	number of non-aromatic conjugated C(sp2)	[324]
H-051	Atom-centered fragments	H attached to alpha-C	[324]
minaaaC	Atom-type E-state indices	Minimum atom-type E-State: ::C:	[430]
minsssN	Atom-type E-state indices	Minimum atom-type E-State: >N-	[430]
MaxsOH	Atom-type E-state indices	Maximum atom-type E-State: -OH	[430]
nLevel7	Chirality descriptors	number of neighboring atoms of the chiral center (level 7)	[324]

The verification of the existence of multicollinearity is a must to ensure that the descriptors are not correlated with one another and vary uniquely with the response variable. Only then can the model be considered statistically significant. To that end, a pairwise correlation plot was performed to check the correlation of the descriptors. Fig IV.6 represents a density histogram of the correlation of descriptors along with the value of the coefficient of determination. All R values are less than 0.75, the threshold value. This diagnosis was also endorsed by the Variance Inflation Factor VIF and tolerance values, as indicated in Table IV.6, vary from 1.24 to 2.08; and from 0.47 to 0.80, which are all less than the threshold value of 3[431], greater than the threshold value of 0.2[431] for VIF and tolerance respectively. Thus, the model is confirmed to be free of potential multicollinearity problems and is statistically significant.

Table IV.6 VIF and tolerance values for the selected descriptors

Descriptor	**VIF**	**Tolerance**
SpDiam_A	1.5725	0.6359
SpMAD_AEA(dm)	2.0892	0.4786
Eig02_AEA (dm)	1.5812	0.6324
nCbH	1.4837	0.6739
nCconj	1.2468	0.8020
H-051	1.4515	0.6889
minaaaC	1.6285	0.6140
minsssN	1.4603	0.6847
MaxsOH	1.3647	0.7327
nLevel7	1.6798	0.5953

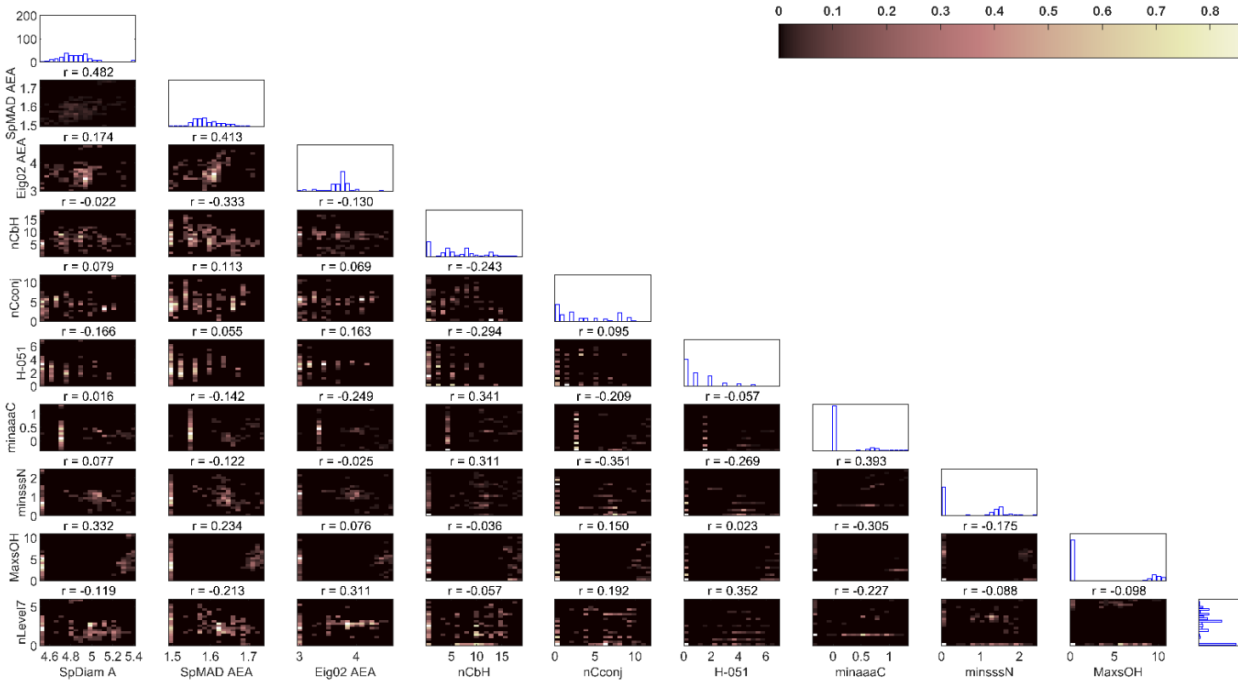
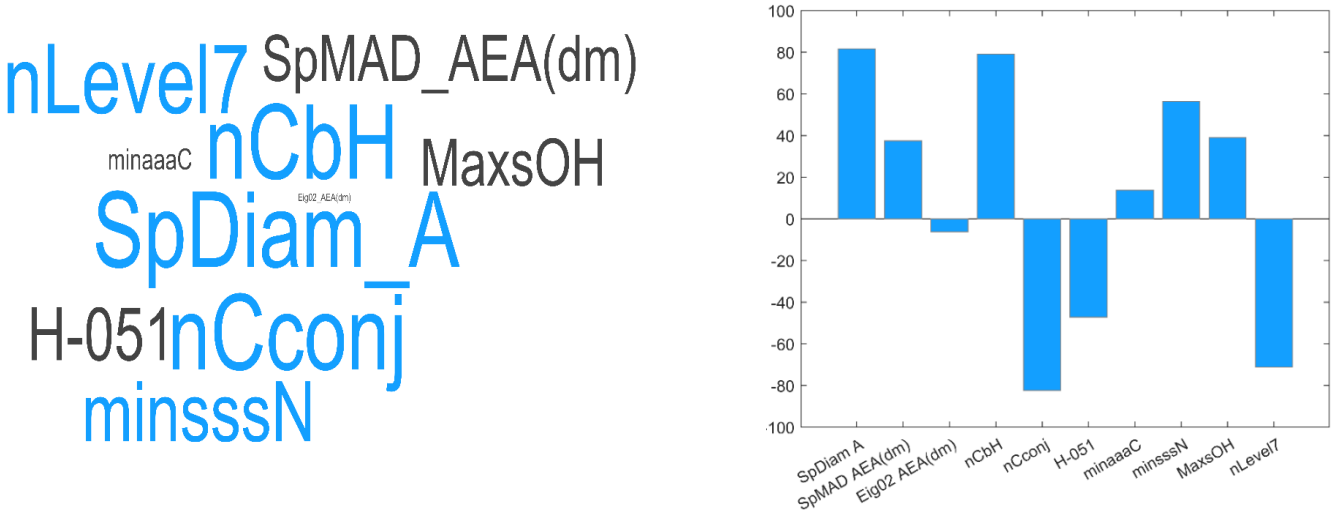


Fig IV.6 Pairwise correlation plot of the selected descriptors

To further assess the mechanistic interpretation of the selected descriptors, the contribution of each descriptor to the inhibitory activity was performed. However, it is known that the SVR method is a black box, making the calculation of the descriptor's importance a complex task, unless other regression methods such as Multiple Linear Regression MLR that uses coefficients, Partial Least Square PLS that uses latent variables, and Artificial Neural Network ANN that uses the weights and bias of the hidden layers[420]. For that, an approach developed by Üstün et al.[432] was followed which allowed the obtainment of the descriptors' relative importance. It consists of using alpha values resulting from the Quadratic Programming QP with the reduced input data matrix (after removing the data that correspond to the zero alphas). A descriptors' profiling vector is obtained after the inner multiplication of the alpha vector and the reduced matrix. The profiling vector was then converted to percentages that sum up to 100% for a better illustration of the results. A word cloud (Fig III.7.a) was constructed to illustrate the absolute importance of each descriptor, where the amplitude of the descriptor's name is proportional to its importance value. SpDiam_A, nCbH, nLevel7, nConj, and minsssN have the greatest effect on the IC_{50}, SpMAD_AEA(dm), H-051, and MaxsOH have a moderate effect on the activity. While minaaac and Eig_AEA(dm) have the least contribution. On the other hand, Fig III.7.b further assesses the word cloud and reveals the type of contribution, whether positive or negative and hence shows which features to increase and which ones to decrease in synthesizing potent inhibitors for the SARS-CoV-2 3CLpro enzyme. There exists a proportional relationship of 81% and 79% between the pIC_{50} and the spectral diameter of the molecule and the number of unsubstituted benzenes respectively (thus inversely proportional with the IC_{50}, meaning the bigger the spectral diameter and the number of unsubstituted benzenes, the smaller the IC_{50}, and thus the more active the inhibitor is). Inversely, the number of nonaromatic conjugated carbon atoms and the number of neighboring atoms of the chiral center at level 7 have a strong negative impact on

the pIC_{50} of -82% and -71% respectively, indicating that the more these structural features are present in an inhibitor, the less active it becomes. In between, the spectral mean absolute deviation weighted by dipole moment, and the minimum electronic state for N and OH have a moderate positive impact on the activity of 37%, 56%, and 38% respectively. On the other hand, conversely, hydrogen atoms attached to alpha carbons were found to negatively influence the inhibiting potential of molecules by a percentage of -47%. Finally, the least influential structural features were the min electronic state of carbon atoms forming two double bonds as well as the eigenvalue weighted by dipole moment with a relative importance of 13% and -6% respectively. As demonstrated, the quantitative interpretation supports and agrees with the qualitative interpretation above.



(a) *(b)*

Fig III.7. Descriptors' relative importance (a) word cloud, (b) histogram

2.2. BR-BPNN model

The BR-BPNN model consisted of 15 neurons using the Bayesian Regularization backpropagation algorithm. Bayesian regularization combats overfitting by penalizing complexity in machine learning models, promoting simpler, better-generalizing solutions. Backpropagation refines model accuracy by adjusting weights based on error gradients, aligning predicted outcomes with actual data[433]. The database was divided into 80% and 20% for the training and the test sets respectively. The visualization of the BR-BPNN model's performance is shown in Fig IV.8, where the predicted pIC_{50} are plotted versus the experimental ones. The plot demonstrates a good accordance between the predicted and the actual (experimental) activity values of the inhibitors of both sets. This is further endorsed by the validation metrics, shown in Table IV.7. The experimental and predicted values by the BR-BPNN model are annexed in Table S-IV.1.

Table IV.7 Statistical validation criteria for the BR-BPNN model

	Internal	**Extern**	**Global**
Systematic Error	Absent	Absent	Absent
R^2	0.8913	0,7220	0,8515

R_0^2	0,8908	0,7050	0,8514
$R_0'^2$	0,8722	0,6961	0,8310
Q^2	0,8907	0,6916	0,8511
$\bar{r}_m^2$	0,8378	0,6442	0,8035
Δr_m^2	0,0750	0,0828	0,0592
CCC	0,9408	0,8470	0,9208
RMSE	0,2682	0,4509	0,3132
SD	0,1918	0,3349	0,2316
MAE	0,1879	0,3056	0,2114

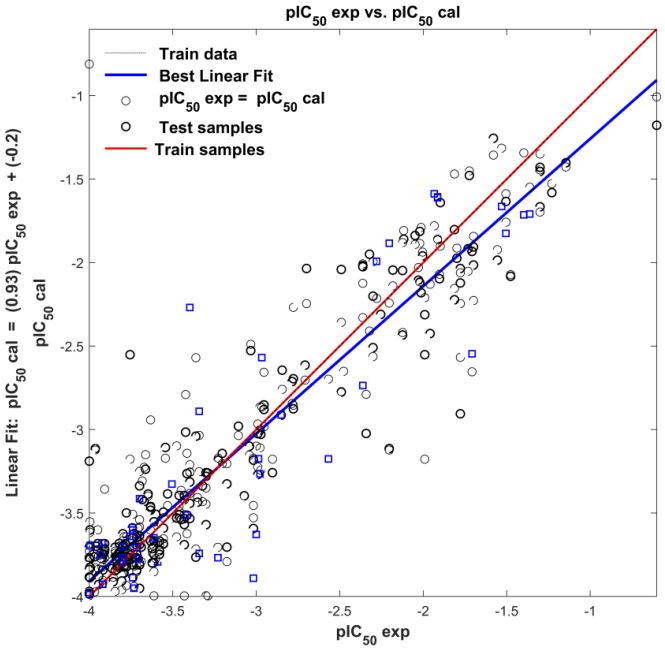


Fig IV.8 BR-BPNN regression plot

- **R^2 (Coefficient of Determination)**

The R^2 value, often known as the coefficient of determination, elucidates the proportion of the variance in the dependent variable (experimental outcomes) that is predictable from the independent variables (descriptors) in the model. An R^2 of 0.8913 internally suggests that approximately 89.13% of the variability in the experimental data is accounted for by the model. Similarly, an external R^2 of 0.7220 implies that the model captures about 72.20% of the variance in unseen data. In essence, a higher R^2 indicates a more accurate prediction of the experimental outcomes.

- **Q^2 (Cross-validation coefficient)**

Q^2 is a crucial metric used in cross-validation, indicating the model's predictive power for new, unseen data points. A global Q^2 of 0.8511 implies that the model can predict nearly 85.11% of the variance in new data, demonstrating its robustness. Q^2 essentially assesses the model's ability to generalize beyond the training dataset. A Q^2 value above 0.5, as seen here, signifies the model's efficacy in making reliable predictions for new observations.

- **$r^{-}m^2$**

This metric indicates the strength and direction of the linear relationship between predicted and observed values. An internal value of 0.8378 suggests a strong positive linear correlation, indicating the model's ability to consistently predict outcomes in line with the experimental data.

- **Δrm^2**

Δrm^2 measures the impact of predictors on the response variable's variance. A value of 0.0750 (less than the threshold value of 0.2, as it should be) internally signifies that the predictors in the model collectively explain 7.50% of the variance in the response variable. This metric helps understand the contribution of descriptors to the overall predictive power of the model.

- **CCC (Concordance Correlation Coefficient)**

CCC evaluates the agreement between predicted and observed values while considering both precision and accuracy. A global CCC of 0.9408 indicates a high level of agreement between the model's predictions and experimental outcomes, emphasizing the model's reliability and precision.

- **RMSE, SD, and MAE**

These metrics quantify the model's prediction errors. A low RMSE (0.2682), SD (0.1918), and MAE (0.1879) imply that, on average, the model's predictions deviate minimally from the actual values. This showcases the model's precision and accuracy in predicting experimental outcomes.

In summary, the BR-BPNN model's robust R^2 and Q^2 values, along with other correlation metrics and error measures, collectively affirm its ability to accurately capture underlying patterns in the data and make reliable predictions for both observed and unseen data points, making it a valuable tool in drug discovery research.

2.3. DA-CNN model

The DA-CNN model's input layer was fed with 10 descriptors as independent variables and the pIC_{50} values as the independent variable. The architecture of the model consisted of 8 layers between convolutional, normalization, pooling and dense layers, in addition to the input and output layers as shown in Fig IV.9. The experimental and predicted values by the DA-CNN model are annexed in Table S-IV.2.

The regression plot of the DA-CNN illustrated in Fig IV.10 indicates that the predicted pIC_{50} are not in good agreement with the experimental values. This shows a potential underperformance of the model. This conclusion is further endorsed by the internal and external validation criteria. The DA-CNN model is found to under-fit with validation metrics failing to meet the acceptable values as shown in Table IV.8.

- **R^2**

Internal R^2 (0.6401): The model explains 64.01% of the variance in the internal dataset. While this demonstrates a moderate fit, it falls below the widely accepted threshold of 0.7, indicating that the model might not capture a significant portion of the underlying variability.

External R^2 (0.4886): External validation R^2 at 48.86% is notably below the threshold, suggesting that the model's predictive power diminishes when applied to unseen data.

Layer (type)	Output Shape	Param #
conv1d_615 (Conv1D)	(None, 8, 64)	256
batch_normalization_615 (BatchNormalization)	(None, 8, 64)	256
conv1d_616 (Conv1D)	(None, 6, 128)	24704
batch_normalization_616 (BatchNormalization)	(None, 6, 128)	512
conv1d_617 (Conv1D)	(None, 4, 256)	98560
batch_normalization_617 (BatchNormalization)	(None, 4, 256)	1024
global_average_pooling1d_205 (GlobalAveragePooling1D)	(None, 256)	0
dense_410 (Dense)	(None, 128)	32896
dropout_205 (Dropout)	(None, 128)	0
dense_411 (Dense)	(None, 1)	129

Fig IV.9 Architecture of the DA-CNN model

Table IV.8 The statistical validation criteria for the DA-CNN model

	Internal	**Extern**	**Global**
Systematic Error	Absent	Absent	Absent
R^2	0,6401	0,4886	0,6029
R_0^2	0,6385	0,4886	0,6020
$R_0'^2$	0,3880	0,2261	0,3057
Q^2	0,6378	0,4859	0,6010
$\bar{r}_m^2$	0,4991	0,3396	0,4592
Δr_m^2	0,2221	0,2651	0,2354
CCC	0,7695	0,6533	0,7427
RMSE	0,4754	0,6378	0,5126
SD	0,3132	0,4065	0,3379
MAE	0,3584	0,4948	0,3861

- **Q^2**

Internal Q^2 (0.6378): Although Q^2 of 63.78% is reasonable, it's just below the 0.7 threshold, indicating that the model captures a substantial portion of variance but struggles to generalize well to new internal data.
External Q^2 (0.4859): External Q^2, falling below 0.5, indicates the model's limitations in predicting new, unseen data, pointing towards a lack of robustness.

- **CCC**

Internal CCC (0.7695): Internally, CCC exceeds the threshold of 0.7, indicating acceptable concordance between predicted and observed values within the internal dataset.
External CCC (0.6533): The external CCC falls below 0.7, signifying a decrease in agreement with new data and indicating limitations in generalization.

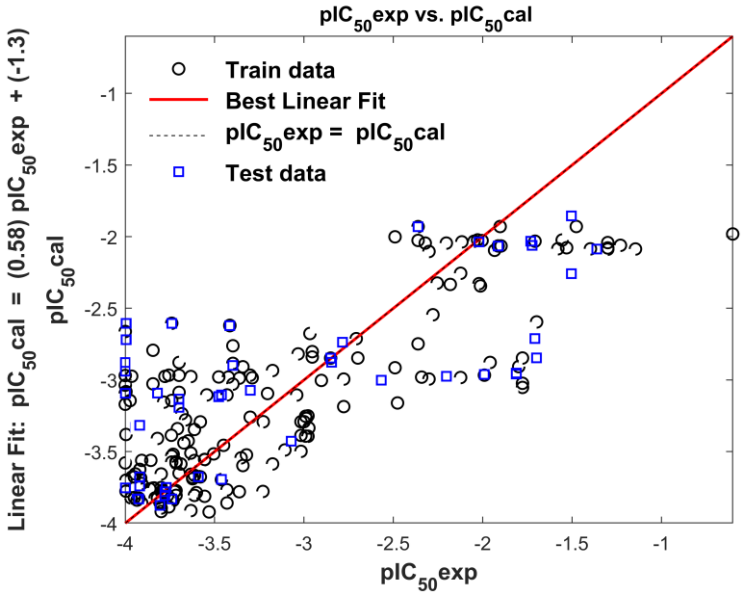


Fig IV.10 DA-CNN regression plot

- **RMSE, SD, and MAE**

RMSE, SD, MAE: These error metrics, exceeding acceptable thresholds, highlight significant prediction errors in both internal and external datasets. These values suggest that the model's predictions deviate substantially from the actual values, pointing towards a lack of accuracy.

The underwhelming performance across various metrics suggests a potential issue with model complexity. The model might be overly intricate, attempting to capture nuances in the data that do not contribute significantly to the underlying patterns. This complexity can lead to overfitting, where the model becomes too tailored to the training data and fails to generalize effectively.

To address this complexity concern, a mid-complex soft computing technique is expected to yield better results than the BR-BP-NN and DA-CNN models. By striking a balance between capturing essential patterns and avoiding unnecessary intricacies, the model can potentially improve its generalization ability and align better with the modeled phenomena.

2.4. DA-SVR model

The algorithm took 10 descriptors as the input and the ordered pIC_{50} values as the response, for 246 molecules divided randomly into 80% for the learning, and 20% for the external test. To obtain the best performance of the hybrid methods, careful tuning of the hyperparameter of the SVR is needed. This task was performed by the DA algorithm: solving the optimization problem of selecting the best combination of C, σ, and ε to obtain the minimal RMSE in the SVR model. Also, the choice of the kernel function was made through the trial-and-error method. The final structure of the SVR model that yielded the best performance is shown in Table IV.9. The full experimented and predicted values by the DA-SVR model are shown in Table IV.10.

Table IV.9 The structure of the DA-SVR model

Parameter	C	Σ	E	Kernel Function
Value	899.35	0.7404	0.010	Gaussian

Table IV.10 DA-SVR experimental and predicted predicted values,train values in pink, and test values in green

pIC50 experimental values	**Predicted by DA-SVR**
-0,602	-0,612079
-1,146	-1,156341
-1,301	-1,311114
-1,301	-1,31096
-1,362	-1,372187
-1,398	-1,332862
-1,477	-1,899722
-1,531	-1,371059
-1,708	-1,718154
-1,708	-1,718321
-1,724	-1,714146
-1,732	-1,741938
-1,778	-1,788034
-1,778	-2,68894
-1,778	-1,788178
-1,799	-1,809031
-1,813	-1,822883

-1,813	-1,823236
-1,903	-1,913383
-1,903	-1,913383
-1,914	-1,918469
-1,934	-1,90943
-1,959	-1,969146
-1,991	-2,001154
-1,991	-2,57151
-2,004	-1,993934
-2,021	-2,03133
-2,117	-2,10706
-2,124	-2,133655
-2,124	-2,134445
-2,182	-2,084179
-2,204	-2,213944
-2,255	-2,264901
-2,279	-2,289325
-2,301	-2,311341
-2,301	-2,310646
-2,322	-2,311882
-2,362	-2,352358
-2,362	-2,352201
-2,362	-2,372157
-2,477	-3,162693
-2,491	-2,501472
-2,568	-2,577551
-2,699	-2,68894
-2,708	-2,697784
-2,778	-2,788152
-2,785	-2,794791
-2,845	-2,534923
-2,851	-2,855166
-2,851	-2,861445
-2,903	-2,91281
-2,954	-2,963744
-2,967	-3,141517
-2,973	-2,983148
-2,978	-2,988358
-2,987	-2,996875
-2,991	-3,001178
-2,991	-3,396561
-3	-3,010015
-3,017	-3,027114

-3,017	-3,027029
-3,017	-3,026756
-3,033	-2,994441
-3,041	-3,558018
-3,072	-3,082435
-3,107	-3,117089
-3,176	-3,679128
-3,176	-3,186271
-3,204	-3,214054
-3,228	-3,027029
-3,29	-3,191138
-3,301	-3,16268
-3,301	-3,291178
-3,301	-3,311281
-3,322	-3,331725
-3,342	-3,331622
-3,342	-3,435142
-3,362	-3,371503
-3,362	-3,352088
-3,398	-3,387716
-3,398	-3,387723
-3,398	-3,162674
-3,398	-3,387604
-3,415	-3,40474
-3,415	-3,405065
-3,415	-3,405065
-3,431	-3,420538
-3,431	-3,440913
-3,431	-3,451812
-3,431	-3,440939
-3,45	-3,315973
-3,462	-3,451812
-3,462	-3,472227
-3,462	-3,451567
-3,477	-3,466545
-3,477	-3,467165
-3,505	-3,494839
-3,531	-3,521146
-3,531	-3,521146
-3,544	-3,53365
-3,556	-3,781705
-3,556	-3,545729
-3,556	-3,56635

-3,58	-3,590353
-3,58	-3,590215
-3,58	-3,569486
-3,591	-3,601133
-3,613	-3,623032
-3,613	-3,602645
-3,613	-3,623015
-3,613	-3,622961
-3,613	-3,602896
-3,623	-3,612723
-3,623	-3,633425
-3,633	-3,623463
-3,633	-3,62307
-3,633	-3,176702
-3,648	-3,638084
-3,672	-3,662179
-3,683	-3,673358
-3,699	-3,180108
-3,699	-3,689437
-3,699	-3,689143
-3,699	-3,688923
-3,699	-3,689016
-3,699	-3,688925
-3,708	-3,698145
-3,708	-3,698145
-3,716	-3,62545
-3,716	-3,706038
-3,716	-3,719975
-3,724	-3,996492
-3,74	-3,749591
-3,74	-3,729825
-3,74	-3,749591
-3,74	-3,729758
-3,74	-3,730008
-3,74	-3,439511
-3,742	-3,732203
-3,756	-3,765948
-3,763	-3,772734
-3,771	-3,761273
-3,778	-3,767952
-3,778	-3,788149
-3,778	-3,78196
-3,778	-3,792781

-3,785	-3,774611
-3,792	-3,781926
-3,792	-3,782402
-3,792	-3,78196
-3,799	-3,78933
-3,799	-3,194608
-3,799	-3,788651
-3,799	-3,788965
-3,799	-3,788651
-3,806	-3,795976
-3,806	-3,796053
-3,82	-3,810095
-3,826	-3,270906
-3,833	-3,822904
-3,833	-3,822728
-3,839	-3,828945
-3,845	-3,163005
-3,845	-3,835279
-3,845	-3,835441
-3,857	-3,846743
-3,863	-3,163513
-3,908	-3,897917
-3,908	-3,89835
-3,908	-3,897706
-3,914	-3,903867
-3,919	-3,9092
-3,919	-3,908951
-3,924	-3,913599
-3,924	-3,533983
-3,934	-3,924177
-3,934	-3,923691
-3,944	-3,445455
-3,944	-3,933634
-3,949	-3,692901
-3,959	-3,948861
-3,963	-3,165711
-3,964	-3,162683
-3,968	-3,957522
-3,973	-3,962965
-3,996	-4,139297
-3,996	-3,9858
-3,996	-3,749591
-3,996	-3,986156

-3,996	-3,986088
-4	-3,220452
-4	-3,989486
-4	-3,989503
-4	-3,990098
-4	-3,990012
-4	-3,990349
-4	-3,989646
-1,23	-1,240318
-1,301	-1,310661
-1,505	-1,515225
-1,505	-1,514936
-1,556	-1,566276
-1,58	-1,570114
-1,699	-1,708759
-1,699	-1,708981
-1,898	-1,797152
-1,914	-1,899977
-2,013	-2,02299
-2,021	-2,010688
-2,029	-2,038843
-2,049	-2,038885
-2,204	-3,162683
-2,342	-2,351938
-2,362	-2,352201
-2,491	-3,162879
-2,778	-2,968708
-2,778	-2,788152
-2,954	-2,963498
-2,978	-2,988178
-3,23	-3,240244
-3,352	-3,168828
-3,423	-3,41285
-3,462	-3,440913
-3,602	-3,592256
-3,633	-3,623041
-3,663	-3,652951
-3,689	-3,679128
-3,699	-3,689143
-3,716	-4,237028
-3,716	-3,726313
-3,732	-3,721902
-3,732	-3,722156

-3,74	-3,729913
-3,756	-2,57151
-3,778	-3,768463
-3,799	-4,420005
-3,82	-3,822728
-3,82	-3,810105
-3,919	-3,909041
-3,919	-3,909022
-3,919	-3,908853
-3,949	-3,939124
-3,996	-3,40215
-4	-3,989832
-4	-3,719975
-4	-3,989503

To visualize the DA-SVR performance, the predicted values of the pIC_{50} were plotted against the experimental values indicating the train set in black and the test set in red, as shown in Fig IV.11, supplementary file I in the supplementary material contains all the observed and the predicted pIC_{50} values and the residuals. Accordingly, response values predicted by the suggested model are in accordance and agreement with the experimental ones, as they, in most, align well with the first bisector. The vector of agreement for both the training and test sets is [a (slope), b (interception y), R (correlation coefficient)] = [0.92, -0.2, 0.9639]. The ideal agreement vector is [1, 0, 1]. This reveals close-to-ideal modeling of all the data by the optimized SVR model. There exist a few molecules whose predicted values do not match their corresponding experimental values. This is considered a normal phenomenon to observe in the regression model as it represents the model's mispredictions but also its generalization ability. In fact, it is a healthy phenomenon that indicates that the model is less likely to overfit the data, and can safely predict new data which is the goal of developing QSAR models. The performance is further judged through the internal and external validation criteria shown in Table IV.11. The internal coefficient of determination R^2 is 0.9365, indicating that the model was able to well learn the nonlinear relation between the 10 selected descriptors and the pIC_{50}. The external R^2 is 0.9078, showing that the pIC_{50} values predicted by the DA-SVR model are in good agreement with the experimental values. This also indicates that the model has a good interpolation and generalization ability, and gives good results when used for a novel set of molecules with an unknown inhibitory activity. It is also worth mentioning that the train and test R^2 are close to each other, signifying that not only did the model learn well, but it also could predict well on unseen data and is also safe from the overfitting problem, where the model overlearns (very high internal R^2) and underperforms on the test data (very low external R^2), and from the underfitting problem where both the internal and external R^2 are low (below the threshold value of R^2= 0.6[321], which was passed by the suggested DA-SVR model). The value of the cross-validation parameter Q^2 is 0.9356 and 0.9059 for the internal and external validation respectively, which is above the threshold value of 0.5[43]. This indicates the model is robust. Also, the value of R^2-Q^2 indicates whether or not the model is

overfitted, and it should be less than 0.3[43]. For the suggested DA-SVR model, the difference between R^2 and Q^2 is 0.0004 for the whole set. This confirms the previous interpretation of the absence of overfitting of the regression graph and R^2 values. Further analysis of the statistical metrics leads us to the discussion of the rm parameters. To safely consider that the predicted activity data calculated by the proposed DA-SVR model is in propinquity with the experimental ones, an evaluation of the rm metric is a must.$\bar{r}m^2$and Δrm^2 measure the average and the difference of the correlation coefficient between the observed and the Leave One Out predicted values of the inhibitors with and without the intercept[321]. This is a tight evaluation metric that indicates that the results obtained by the model were not due to chance. The developed DA-SVR model has $\bar{r}m^2$ and Δrm^2 values of 0.8947 and 0.04 for the whole dataset, respectively, which satisfy the acceptance criteria values of $\bar{r}m^2>0.5$ and $\Delta rm^2<0.2$[321]. Also, the Concordance Correlation Coefficient CCC was used to evaluate the model's performance. The global CCC value is 0.9632, which indicates that the developed model is reliable. Furthermore, an evaluation based on the error metrics was carried out. This included the Root Mean Squared Error RMSE, the Standard Deviation SD, and the Mean Absolute Error MAE. Their respective global values are 0.2166, 0.1988, and 0.0870 indicating that the model is precise and further sustains the powerfulness of the DA-SVR in predicting the IC_{50} of potential inhibitors.

In summary, the discussion of the internal and external validation criteria confirms that the proposed DA-SVR model is robust, reliable, and precise in predicting the activity of the potential inhibitors.

Table IV.11 Statistical validation criteria for the DA-SVR model

	Internal	**Extern**	**Global**
Systematic Error	Absent	Absent	Absent
$\mathbf{R^2}$	0.9365	0.9078	0.9291
$\mathbf{R_0^2}$	0.9365	0.9064	0.9291
$\mathbf{R_0'^2}$	0.9323	0.9049	0.9253
$\mathbf{Q^2}$	0.9356	0.9059	0.9287
$\mathbf{\bar{r}_m^2}$	0.8900	0.8760	0.8947
$\mathbf{\Delta r_m^2}$	0.0455	0.0069	0.0496
CCC	0.9665	0.9527	0.9632
RMSE	0.1964	0.2838	0.2166
SD	0.1799	0.2615	0.1988
MAE	0.0797	0.1164	0.0870

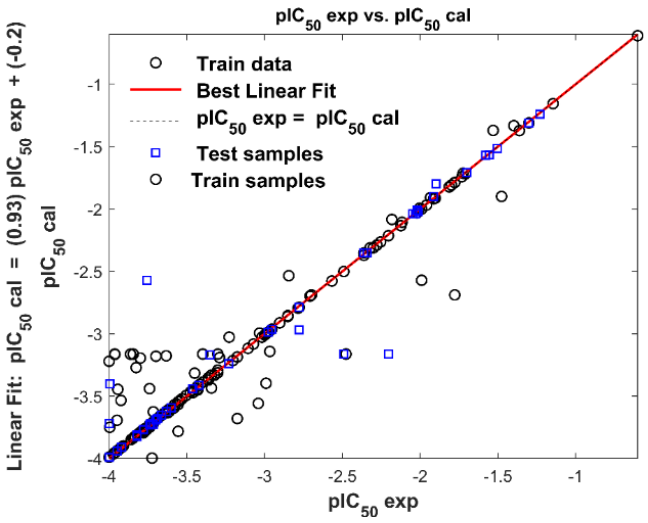


Fig IV.11 DA-SVR regression plot

2.5. Models' comparison

The BR-BPNN model's effectiveness can be attributed to Bayesian regularization, which mitigates overfitting by promoting simpler solutions. This regularization method prevents the model from becoming overly complex, enhancing its ability to generalize to unseen data. Additionally, the backpropagation algorithm refines accuracy by adjusting weights based on error gradients, aligning predictions with actual data.

The DA-CNN model's potential overfitting issues arise from its complexity. Deep architectures, if not appropriately managed, tend to memorize the training data rather than learning underlying patterns. This memorization can lead to poor generalization, especially with small datasets, as evidenced by the model's struggles with unseen data.

The DA-SVR model's superior performance is a result of meticulous hyperparameter tuning using the Dragonfly Algorithm. By optimizing parameters such as C, σ, and ε, crucial components of SVR, the model strikes a balance between complexity and the ability to generalize. This precise tuning ensures the model captures essential patterns while avoiding overfitting, leading to accurate predictions and robust generalization. DA-SVR, guided by the Dragonfly Algorithm, represents a compromise between complexity and the ability to generalize, making it particularly effective, especially with small datasets. It outperforms neural network algorithms, especially in scenarios involving noisy and limited data, where overfitting is a significant concern.

Among the soft computing models, the BR-BPNN model stands out due to its ability to strike a balance between complexity and simplicity. Bayesian regularization prevents overfitting, enhancing its generalizability. However, further exploration of hyperparameters could enhance its predictive power. In contrast, the DA-CNN model's challenges primarily stem from its complexity[417]. The deep architecture, if not managed carefully, tends to overfit, especially in scenarios with limited data. Thus, finding the right balance between capturing

intricate patterns and avoiding unnecessary complexities is crucial for its improvement. Finally, the DA-SVR model excels due to its precise parameter optimization using the Dragonfly Algorithm and the compromise it represents between complexity and generalizability. By avoiding overfitting and capturing essential patterns, the model accurately predicts outcomes even in challenging scenarios with small and noisy datasets. Its high level of agreement with experimental data showcases its reliability, making it particularly valuable for complex modeling tasks. Therefore, solely the DA-SVR model is considered for further analyses and discussion.

2.6. Applicability domain

The AD was represented with the William plot in Fig IV.12, where the standard residual is plotted versus the leverage. The two horizontal lines represent the residual cutoff values -3,3, and the vertical line is the critical leverage h* of 0.14.98.78% of the inhibitors fell before the critical h*, and only 3 molecules (1.21%) exceeded the critical leverage, thus being potential structurally influential molecules. Considering the response outliers, only 4% of the 246 molecules fell outside the cutoff values, while the rest 236 inhibitors(96% of the dataset) were covered by the domain [-3,3], mostly aligning with the zero standard residual value. Thus, the model had a good applicability domain that covered the vast majority of the training and test sets, and the analysis of the AD confirmed the interpretation of the model's performance above. Accordingly, the model has been demonstrated to be well-performing, robust, statistically valid, and powerful in predicting the pIC_{50} activity of the selected inhibitors, and can safely be used in predicting the inhibitory activity of novel molecules against 3CLpro.

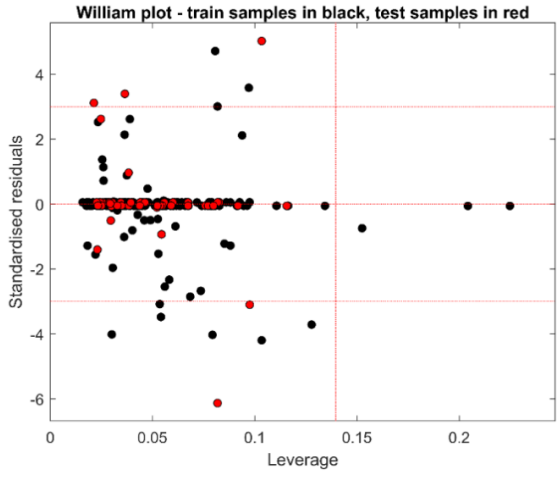


Fig IV.12 Applicability domain of the DA-SVR model

2.7. Comparison with literature

To further assess the quality of the developed DA-SVR model, a comparison with other recently reported models (2021-2022) that target the same enzyme was carried out. The comparison was based on one correlation metric R^2 and one error metric RMSE, illustrated in Fig IV.13. To make a more purposeful comparison, external metrics were chosen, because a

good model is a model that performs well on unseen data. The comparison was illustrated by a grouped bar plot that shows the R^2 and RMSE of the QSAR models reported in the studies [368],[371],[364],[365], and [267] respectively. The DA-SVR model proposed by this study outperforms the rest of the models in terms of agreement with the predicted and experimental values and in terms of precision in predictions. Henceforth, the hybrid DA-SVR model is advantageous and performing, and can be used for the prediction of unknown activities for potential inhibitors.

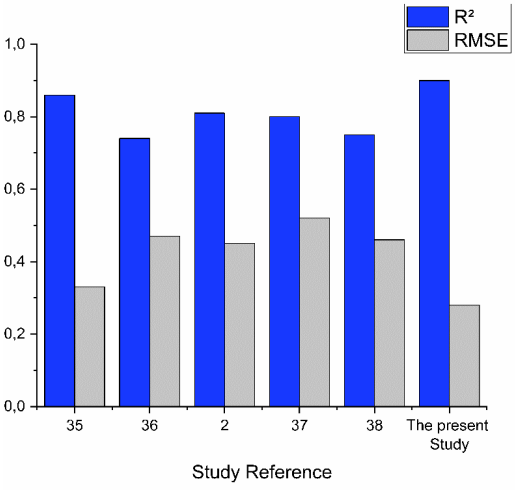


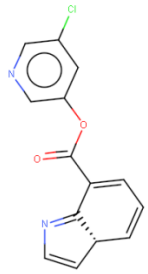
Fig IV.13R^2 and RMSE bar plot for some literature-reported models and the present DA-SVR model

2.8. ADMET properties

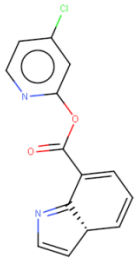
The evaluation of the ADMET properties of the 246 molecules was assessed. To assess their drug likeliness, the Lipinski Rule of Five was considered. Molecules violating 2 or more rules were discarded. The filtrations resulted in 24 molecules with good drug-like properties and thus were kept for further analysis. According to the threshold values on the webservers used, the molecules fall well in the range of optimal values and satisfy the rules. The Lipinski Rule of five assures that the molecules have good absorption when there are no more than 5 hydrogen bond donors, 10 hydrogen bond acceptors, 500 for molecular weights, and 5 for solubility examined by the water/octane partition coefficient. The gastrointestinal absorption parameter GIA shows that the molecules have high absorption except for one. This might be due to the sulfonic acid functional group that is present solely in this molecule and absent in the rest. This explains why it is the only molecule with a low GIA. It is worth mentioning that this molecule (number 6) did not violate any rule of Lipinski. However, the replacement of the OH group in the sulfonic acid by an amide group (NR2) to produce a sulfonamide could result in a better GIA. The Pfizer rule indicates possible toxicity. It is violated when the LogP is greater than 3 and the topological polar surface area TPSA is less than 70. Only 6 molecules failed to have satisfactory LogP and TPSA values. These molecules (1, 2, 4, 20, 21, and 22, as shown in Fig IV.14), out of 8 chlorine-containing molecules, all contain chloropyridine or chlorobenzene. Chlorine and benzene do not contribute to the total

topological polar surface area but have higher LogP values, unlike oxygen and nitrogen, which explains the relatively low TPSA and high LogP of these compounds. GSK and the Golden Triangle Rules evaluate the general ADMET profile as they take into consideration molecular weight, LogP, and LogD. The latter evaluates the lipophilicity of a molecule in the physiological pH of the blood. These are crucial parameters in determining the permeability of the drug-like molecules as well as their hydrophobic bindings with various proteins including those involved in metabolism and excretion. None of the 24 drug-like molecules violated the GSK and the Golden Triangle Rules, indicating that they have a good ADMET profile. Moreover, the molecules fully respected the Ghose, Veber, and Muegge Rules, which take into consideration the molecular weights, various lipophilicity parameters, molar refractivity, number of atoms, number of rotatable bonds, number of rings, number of heteroatoms, number of carbon atoms, TPSA and hydrogen bond donors and acceptors. These parameters reflect the toxicity of the molecules as well as their pharmacological properties and thus their ADMET profiles. Another parameter is the Synthesis Accessibility Score SA score which indicates how easily a molecule can be synthesized on a scale from 0 to 10. Compounds with a SA score of less than 5 are considered easily synthesized. According to the results, all 24 drug-like molecules can easily be synthesized as they have an acceptable structural complexity and ordinary functional groups.

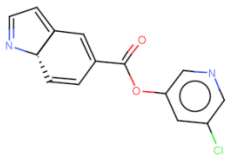
The overall evaluation of the selected 24 molecules reveals an excellent drugability and ADMET profile as they fall in agreement with the various evaluation rules. Therefore, they can safely be considered to possess good pharmacological properties and used in further analysis.



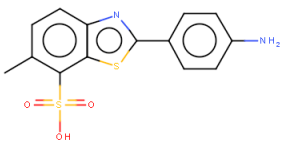
Molecule 1



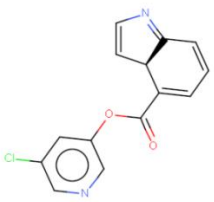
Molecule 2



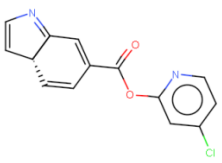
Molecule 4



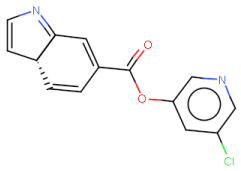
Molecule 6



Molecule 20



Molecule 21



Molecule 22

Fig IV.14 Structure of molecules with low GIA (molecule 6) and violations of the Pfizer Rule

Table IV.12 Major rules evaluation of the 24 molecules

Molecule Number	Lipinski Rule	Pfizer Rule	GSK Rule	Golden Triangle	Ghose Rule	Veber Rule	Muegge Rule	Leadlikeness Rule
1	Accepted	Rejected	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted
2	Accepted	Rejected	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted
3	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted
4	Accepted	Rejected	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted
5	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted
6	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted
7	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted
8	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted
9	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted
10	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted
11	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted
12	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted
13	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted
14	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted
15	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted
16	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted
17	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted
18	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted

19	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted
20	Accepted	Rejected	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted
21	Accepted	Rejected	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted
22	Accepted	Rejected	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted
23	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted
24	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted	Accepted

2.9. Molecular docking

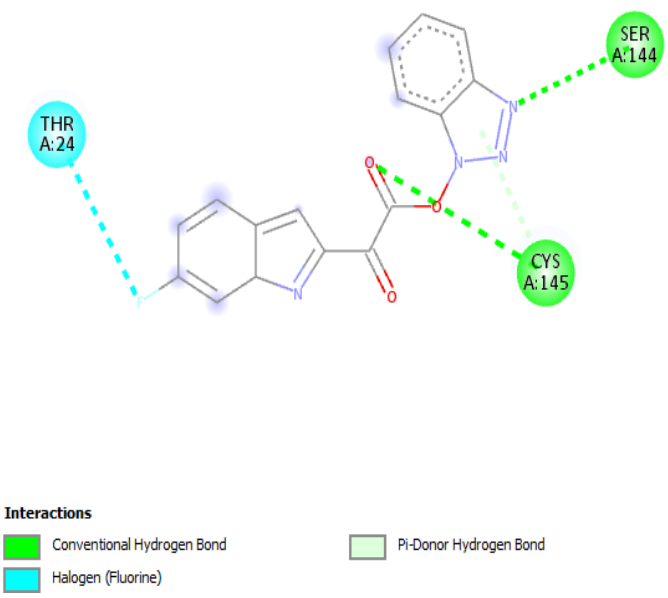
The 24 molecules displayed a good affinity ranging from -7.8 kcal/mol to -5.6 kcal/mol. The top two molecules with the best (smallest) affinity are molecules 19 and 15 exhibiting affinity energy values of -7.8 and -7.1 kcal/mol respectively. While the least affine molecules are 13 and 7 with an affinity of -5.6 kcal/mol for both. The visualization of these 4 molecules' interactions with the active site of the 3CLpro enzyme is shown in Fig IV.15, while those of the total 24 molecules can be found in Table S-IV-5 in the annex. Molecule 19 exhibited 3 types of interaction: conventional hydrogen, Pi-donor hydrogen, and alkyl. Conventional hydrogen bonds HB happen when hydrogen forming a covalent bond with an electronegative element such as oxygen and nitrogen forms another weaker bond to another electronegative atom[434]. Molecule 19 formed 3 conventional HB at 3 different sites: N in the triazole group with Ser144, O in the ester group with Cys145, and N in the 1H-isoindole group with Ser46. Their lengths are 2.81, 3.51, and 3.17 Å respectively. HB are key factors in determining the specificity of the ligand binding[435] as they are considered a crucial factor in the recognition of the ligand by the target enzyme[436]. Moreover, an HB contributes up to 2kcal/mol to bonds' strength, depending on the polarity of the surrounding environment, solvent-exposed interactions are 1.2kcal/mol weaker than those in the hydrophobic interior of the protein[437]. The fact that the conventional HB were present in all the interactions of the 24 molecules indicates that they have specificity with and can be well recognized by the 3CLpro enzyme, which is one of the reasons that even the least performing molecule still had a relatively good affinity (-5.64 kcal/mol). Pi-donor hydrogen bonds consist of polar hydrogen of an electronegative group and an aromatic group. Since triazole is a pi-excessive aromatic group, it forms a pi-polar hydrogen bond with the Cys145 as it contains polar hydrogen atoms from the NH2, OH, and SH functional groups[438].

Molecule 15, the second-best performing drug-like molecule, also showed conventional HB: N in the triazole group with Ser144, O in the ester group with Cys145 respectively, and pi-donor HB between sulfur in Cys145 and triazole. In addition to the Florine bond with Thr24 and the Florine substituent in the 1H-isoindole fragment. Florine is widely present in drugs (50% of profitable drugs contain Fluorine)[439]. Fluorine forms halogen bonds with amino acid residues in the active site of the enzyme (Thr24 in this case) which are very common in drug-target interactions[408]. Molecule 15 differs from 19 in the presence of Fluorine

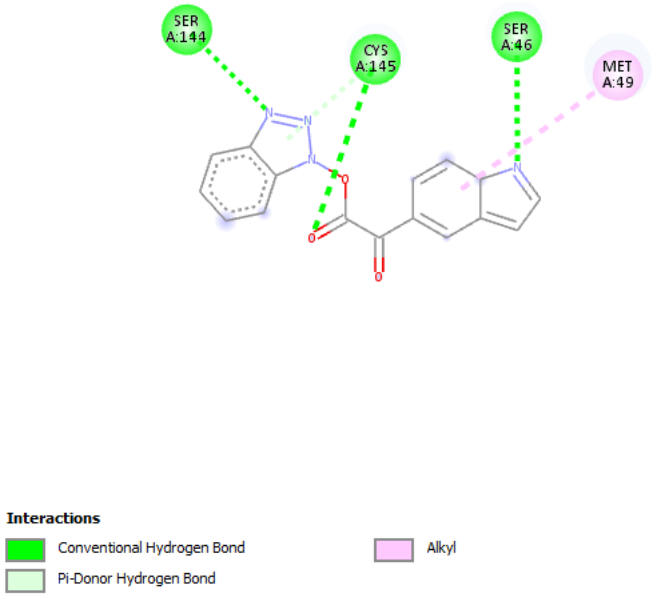
attached with 1H-isoindole. However, this structural change caused a conformational change leading the Ser46 and Met49 to be far from the 1H-isoindole fragment and thus the absence of the third conventional HB and alkyl bond, slightly lowering the affinity of the ligand with the enzyme. Precisely, the addition of Fluorine to the 1H-isoindole seems to have a greater attraction with Thr24 (Thr was observed to be very reactive with aromatic Fluorine[439]) than with Ser49 and Met49 since the halogen bond between Fluorine and Oxygen in Thr24 is attractive and stabilizing by nature[440]. Therefore, the attractiveness of this non-covalent bond can induce a conformational change in ligands that overshadow conventional HBs and alkyl bonds. If the benzo-triazole fragment had a Fluorine substituent, it is presumed to be more affine and active. The two least performing molecules are 13 and 7 with an affinity of -5.6kcal/mol. Molecule 13 formed one pi-sulfur bond of 5.17 Å between Cys145 and the pi-electron cloud in the double bonds of pyridine, and two other HBs between Thr45 and oxygen, and Thr26 with nitrogen with bond length of 3.23Å and 3.11Å respectively. Molecule 7 interacted with 6 bonds of 4 types: HBs with Leu141 and Cys145 with lengths of 2.66Å and 3.31 Å, pi-sigma and pi-donor bonds with the methyl group and the primary amine group pi-electronpairin the conjugation of the double bonds of chlorobenzene acting as the electron rich element, with bond lengths of 3.99Å and 4.19Å respectively, and pi-alkyl bond between the sulfur of Cys145 and the diaminopyrimidine functional group (4.19Å) along with another pi-donor bond (3.15Å) with Gly143, precisely with the primary amine present. It is noticed that chlorine present in both structures, despite its susceptibility to react with residues in the active site, showed no interactions whatsoever with aminoacids in the active site of 3CLpro. This is due to the considerable distance, in that specific (best) pose, between chlorine and the active site residues, as it was exposed to the solvent and not directed to the interior of the pocket. To further prove this claim, the second-best poses of molecules 7 and 13 were examined. These poses scored -5.4 kcal/mol for both molecules, within which chlorine was close the active site's residues and oriented inwards. Chlorine in molecule 7 formed 2 alkyl bonds with carbon atoms in Leu27 andCys145 of length 4.31Å and 3.35Å respectively. Molecule 13 also showed 2 alkyl bonds between chlorine and Leu27 and Cys145 (4.24Å and 4.20 Å) respectively, and another pi-alkyl bond with the electron deficient ring in His41 (4.42Å). However, these bonds formed by chlorine caused the disappearance of other bonds found in the first best scoring pose for both molecules, such as pi-donor HBs and pi-sigma bonds, and longer HBs and pi-sulfur bonds (5.85Å) in molecule 7. Henceforth, the best scoring poses for molecules 7 and 13 included no chlorine-involved bonds. Regarding the strong interactions made by fluorine in molecule 15 and the absence of chlorine interaction in all chlorine-containing molecules, it can be deduced that fluorine is a more potent halogen than chlorine, indicating that when designing an analogue, it is desirable to consider fluorine over chlorine.

The docking results are in strong agreement with the above-mentioned descriptors' analysis of the QSAR study. The descriptor SpDiam_A (spectral diameter from the adjacency matrix) was reflected by the repetitive presence of pi-bonds of different types as SpDiam_A suggests that the more negative the HOMO energy, the more reactive the inhibitor is (since the negative of the HOMO energy corresponds to the first ionization energy of a molecule, according to the Koopman's theory). Pi-bonds of the studied molecules, molecule 19 for

example, consisted of electron-rich species (high HOMO energy) and electron-deficient species, where the electron-rich species was always a functional group from the studied molecules, more often the triazole group. The descriptor minsssN reflects the Minimum atom-type E-State of nitrogen atoms with 3 single bonds; and as suggested by the QSAR findings above, nitrogen presence contributes to the formation of HBs, which are a key factor in drug-target recognition but also its stabilization. The descriptor maxsOH, representing the number of terminal/radical alcohols, is present in 4 molecules: 14,18, 23, and 24, with an affinity of -5.8,-6, -6.7, and -5.8 kcal/mol respectively, and is related to HBs interestingly formed between the oxygen and the hydrogen of the alcohol (in the case of molecule 24) mainly with Thr45 and Ser46 among others. The descriptor nlevel(7) describes the number of neighboring atoms at shell 7 from the chiral center. As interpreted in descriptors analysis, the docking results confirm that, for similar molecules (molecules belonging to the same cluster), the more neighboring atoms, the less potent the inhibitor is. This is illustrated in molecules 19 and 15, where they are similar to each other, the best-affinity molecule (19) has only 1 neighboring atom at the 7th shell of the chiral center, while the second-best-affinity molecule (15) has 2. This can be seen through the rest of the molecules as well. Thus, this descriptor, confirmed by the docking results, suggests that in the chiral pair of molecules (19,15), molecule 19 is the best recognized by and affine to the 3CLpro enzyme.



(a)



(b)

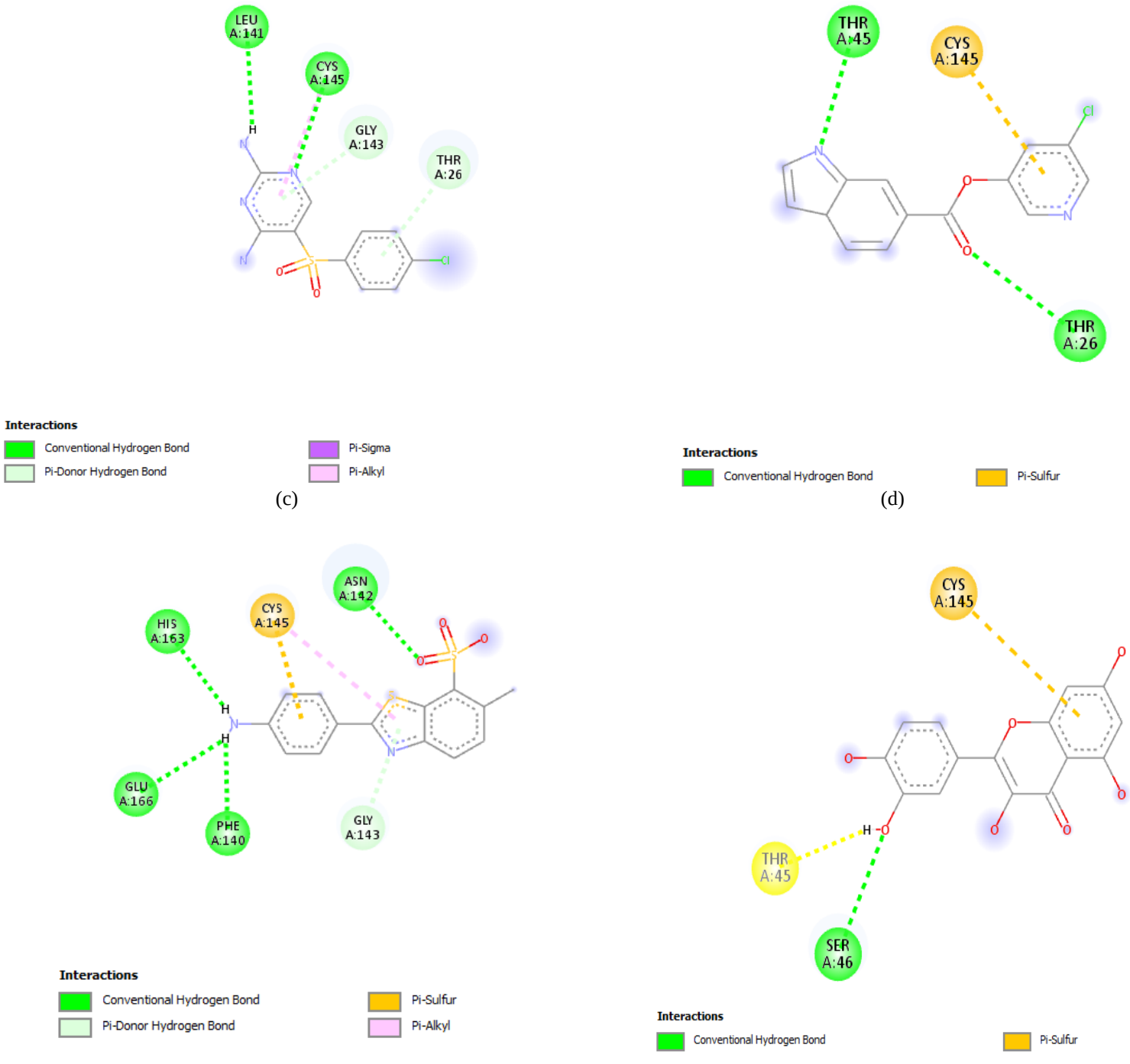
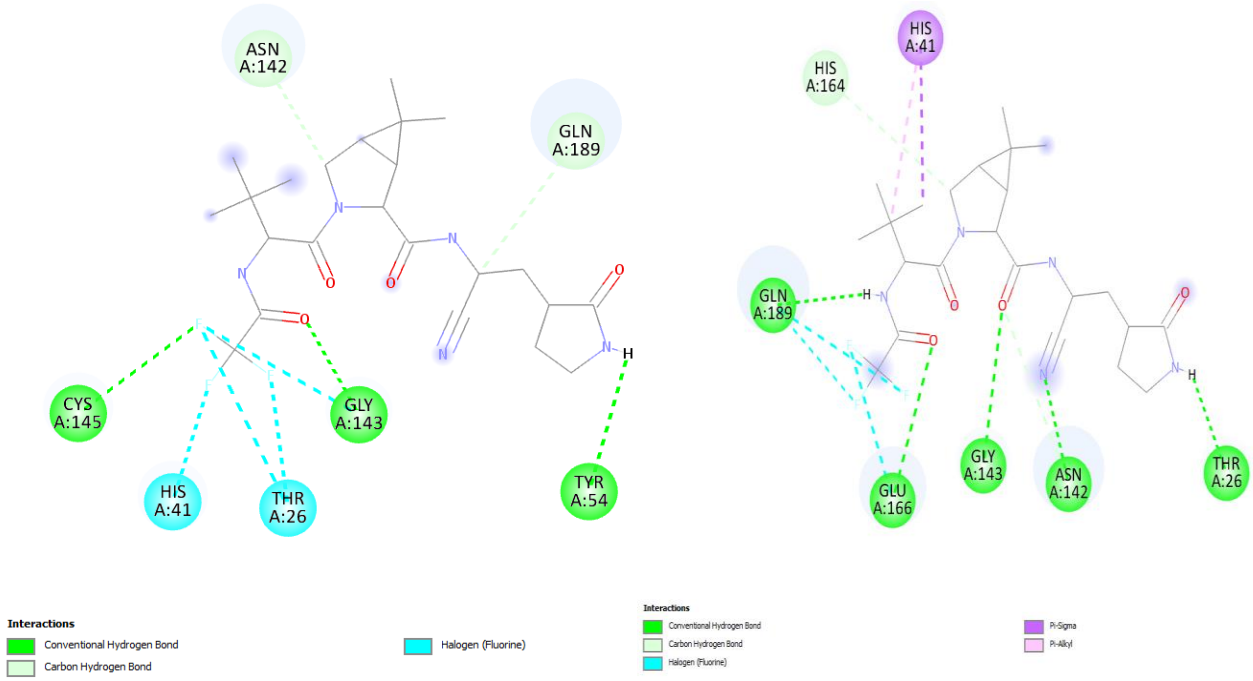


Fig IV.15 Interactions of molecules (a) 15, (b): 19, (c): 7, (d): 13, (e): 18, (f): 24

The reference molecule is the Pfizer molecule PF-07321332 illustrated in Fig IV.16, which is currently in advanced clinical trials. It was docked using the same procedure. The molecule scored affinity energy of -8.8 kcal/mol, with 3 conventional HBs (Cys145 with fluorine, Gly143 with oxygen in the amide group, and Tyr54 with the nitrogen hydrogen of the pyrrolidone of bond lengths of 3.52 Å, 2.87 Å, and 2.64 Å), 3 halogen bonds with the 3 fluorine atoms both with the oxygen present in Thr26 and nitrogen of His41 of lengths of 3.19 Å, 3.86 Å, and 3.03 Å respectively, and 2 carbon HBs between ASN142 carbon the tertiary

nitrogen ring (3.32 Å) and between GLN189 and nitrile carbon (3.40 Å). The second-best-performing pose, with an affinity energy of -7.98 kcal/mol, showed more interactions, notably the nitrile warhead that was involved in a conventional HB with ASN142 (3.13 Å) in addition to new pi-sigma and pi-alkyl bonds. Molecule 19 and PF-07321332 have an affinity energy difference of -1 kcal/mol which is evidence that molecule 19 could be a good alternative for PF-07321332. This difference could be due to the presence of several amide groups whose oxygen and nitrogen are involved in HBs, and the three fluorine atoms in contrary to none/one in molecules 19 and 15.

In summary, HBs and pi-bonds of different types were the most frequent bonds between the 24 inhibitors and the active site residues of the 3CLpro enzyme. Moreover, the QSAR results and docking results fell in agreement with each other, confirming their powerfulness and correctness. Considering the QSAR results, the IC_{50} values, the ADMET properties, and the docking results, it is safe to consider molecules 19 and 15 as potent inhibitors and consider them for further analysis and in wet lab synthesis since they were well described by the molecular descriptors, well predicted by the DA-SVR model, fully accepted by the Lipinski Rule, Pfizer Rule, GSK Rule, and Golden Triangle and thus good pharmacokinetic and pharmacodynamics properties, and they showed an excellent affinity with the binding site of the enzyme.



(a) (b)

Fig IV.16 Interactions of PF-07321332 (a) best pose, (b): second best pose

3. Conclusion

In the work on ML300 analogues, three models were built to predict the inhibitory activity of potentially therapeutic molecules to contribute to the acceleration of drug discovery by the mean of computational chemistry. The results indicate that the relation between the IC_{50} and

the selected descriptors is nonlinear and thus needs a nonlinear approach to model it. Among the three algorithms, the PSO-SVR model performed best with the lowest errors and highest regression coefficients. The PSO-SVR model obtained was demonstrated to be a robust, statistically and biologically meaningful model capable of predicting the activity of unknown potential inhibitors. The obtained results also revealed the key descriptors to maximize the activity of the selected molecules, among which the structural complementarity characteristics in terms of the moment of inertia, rotatable bonds, and weighted paths were found proportional to the activity. On the other hand, a very complex hyper-volume of reactivity may hinder the binding of the molecule and the active site. Further docking and virtual screening studies are being investigated based on the findings of this article. The results obtained in this article have promising aspects in designing a potential drug against COVID-19.

The global aim of modeling heterogeneous inhibitors was to identify potential inhibitors of the 3CLpro enzyme by implementing a ligand-based technique (QSAR) and a structure-based technique (molecular docking). Thus, this study consisted of developing and interpreting correlative and mechanistic QSAR models to predict the half maximum inhibitory activity IC_{50} of potential inhibitors of the 3CLpro enzyme of the SARS-CoV-2 virus, considering ADMET properties as a filtering procedure to select drug-like molecules, and elaborating a molecular docking study of the selected inhibitors to understand the nature of their interacting modes with the enzyme. The mechanistic interpretation had the goal of exploring the biological meaning of the developed model and decoding the underlying structure-activity relationships SARs. Thus, the selection of pertinent descriptors was carried out using SVR and revealed 10 descriptors ranging from adjacency matrix, functional groups, electro-topological, and chirality-derived descriptors. The qualitative and quantitative interpretations agreed with each other and revealed that the larger the mean of the neighboring atoms and the less the HOMO is, the better the inhibitor, with a relative importance of 81%, along with the number of unsubstituted benzene carbons, that impacted the powerfulness of the inhibitors of 79%. Conversely, nonaromatic conjugated carbons and a large number of the atoms 7 bonds distant from the chiral center were found to hinder the binding of the inhibitors with the 3CLpro enzyme of -82% and -71%. The integration of BR-BPNN, DA-CNN, and DA-SVR has propelled our research to new heights, culminating in the creation of the DA-SVR model. With good precision (R^2: 0.92, Q^2: 0.92, RMSE: 0.21), this model exemplifies the capabilities of integrated Soft Computing techniques. Thus, the model was confirmed to be statistically valid, unbiased, and robust, with 98% of the data contained in the applicability domain. The model was able to outperform many other models recently reported in the literature. The ADMET properties indicated that 24 of the 246 molecules had good ADMET profiles and drug-like properties, indicating the possibility of oral administration as drugs. Their docking results suggested that molecules 19 and 15, sharing the 1H-isoindole and triazole fragments showed the best affinity energy of -7.8 and 7.5 kcal/mol, showing a close energy affinity with the reference molecule with affinity energy of -8.8 kcal/mol, with conventional and pi-donor hydrogen bonds, halogen and alkyl bonds, confirming the QSAR results that the spectral diameter, mainly reflected by the triazole functional group and the presence of single-bonded

nitrogen atoms and alcohol groups as shown in the interactions have a good impact on the inhibition of the enzyme.

Based on all the above results, molecules 19 and 15 are potentially potent inhibitors of the 3CLpro with low IC_{50} values, good ADMET properties, and a strong affinity with the enzyme. Thus, they can be used for further screening of larger inhibitor libraries, analogs generation, and QSAR-based synthesis studies to undergo wet-lab experiments and contribute to the discovery of potential leading compounds against the 3CLpro enzyme of the SARS-CoV-2.

Conclusion

General Conclusion

In the annals of scientific inquiry, this thesis is an opus of great significance, representing an expansive odyssey into the intricate realms of Computer-Aided Drug Design. Within these pages, a rich tapestry of methodologies and computational innovations has been woven, promising to reshape the very contours of pharmaceutical discovery. At its heart, this study was a relentless pursuit of understanding, a quest to decipher the cryptic language of molecular interactions and harness this knowledge for the greater good.

The foundational objective of this thesis was nothing less than monumental: to unravel the complex dance between molecular structures and their inhibitory potential, focusing particularly on the 3CLpro enzyme of the notorious SARS-CoV-2 virus. In this endeavor, advanced computational techniques, spanning the spectrum from ligand-based (QSAR) analyses to sophisticated structure-based (molecular docking) simulations, were employed. Through meticulous experimentation, each algorithm and model was meticulously refined transforming raw data into nuanced insights and theoretical concepts into actionable intelligence.

Central to the success of this research were the predictive models, each a testament to the fusion of cutting-edge technology and unwavering dedication. The PSO-SVR model, crowned with laurels, emerged as a beacon of accuracy and reliability, demonstrating an unprecedented capacity to predict the inhibitory activities of a set of chemical compounds. These models not only highlighted the computational prowess at our disposal but also hinted at a future where predictive algorithms become indispensable tools in the arsenal of drug developers.

However, the true marvel of this endeavor lies not merely in its computational achievements but in the deep understanding of molecular nuances it offers. Through exhaustive analysis, a symphony of descriptors was identified, each note contributing to the harmonious melody of inhibitory efficacy. Structural complementarity characteristics, akin to the intricate brushstrokes of a masterful painting, intertwined with specific functional groups, creating a canvas upon which the fate of potential therapeutics was written. This deep comprehension of molecular intricacies not only enriches our theoretical repository but also charts a course for the deliberate design of novel, potent drug candidates.

Delving further into the labyrinthine world of mechanistic interpretation, this research probed the very essence of biological significance. It dissected the core of structure-activity relationships (SARs), unraveling the enigmatic language through which molecules communicate their inhibitory potential. This profound decoding not only elucidates the mysteries of molecular interactions but also presents a roadmap for the future, where drugs are not merely discovered but meticulously crafted from a deep understanding of their biological relevance.

Amidst this scientific symphony, two luminaries emerge: molecules 19 and 15. In their molecular structures lies the promise of a brighter tomorrow. Their exceptional inhibitory properties, combined with favorable ADMET profiles and a robust affinity for the 3CLpro

enzyme, herald a new dawn in the fight against viral infections. These compounds, symbolic of human ingenuity and scientific brilliance, beckon forth a future where diseases are not merely treated but conquered.

In essence, this thesis transcends the boundaries of conventional scientific inquiry. It is not merely a culmination of experiments and analyses; it is a manifesto of possibilities, a testament to the indomitable spirit of exploration. As we gaze upon the horizon of pharmaceutical discovery, this research stands as a lighthouse, illuminating the path ahead Its methodologies, models, and compounds do not merely constitute a chapter in the annals of science; they mark a paradigm shift, inviting us into an era where diseases are not feared but understood, and where today's remedies become the foundations of a healthier and brighter future for all of humanity.

The obtained results and findings open the horizon for many perspectives in finding a therapeutic molecule for COVID-19:

- *In vitro* and *in vivo* evaluation of the suggested molecules
- Analogues generation based on the suggested molecules and on the suggested structural features.
- QSAR-based synthesis and inverse QSAR for the suggested molecules
- Use of other AI techniques such as GNN, and other metaheuristic optimization algorithms such as ABC, GWO and multiple hybridizations.
- Consideration of other potential targets such as Spike protein and RdRp enzyme, and the dual inhibition synergy possibility.

References

[1] B. Hu, H. Guo, P. Zhou, Z.-L. Shi, Characteristics of SARS-CoV-2 and COVID-19, Nat. Rev. Microbiol. 19 (2021) 141–154. https://doi.org/10.1038/s41579-020-00459-7.

[2] J. Wise, Covid-19: Global response was too slow and leadership absent, report finds, BMJ. 373 (2021) n1234. https://doi.org/10.1136/bmj.n1234.

[3] S.N. Klouche-Djedid, J. Shah, M. Khodor, S.E.O. Kacimi, S.M.S. Islam, H. Aiash, Algeria's response to COVID-19: an ongoing journey., Lancet. Respir. Med. 9 (2021) 449. https://doi.org/10.1016/S2213-2600(21)00083-7.

[4] I. ZATLA, L. BOUBLENZA, The COVID-19 outbreak inception in Algeria: What Happened?, (n.d.).

[5] M. Lounis, A Descriptive Study of the Current Situation of COVID-19 in Algeria, Electron. J. Gen. Med. 17 (2020) 1–4. https://doi.org/10.29333/ejgm/8287.

[6] T. Rouabah, A. Tounsi, N.E. Belaloui, A mathematical epidemic model using genetic fitting algorithm with cross-validation and application to early dynamics of COVID-19 in Algeria, 2020.

[7] M. Ababsa, H. Aouissi, Current state of the Coronavirus (Covid-19) in Algeria, (2020).

[8] A. Kada, K. Bouyoucef, K. Sahraoui, IMPACT OF HYDROXYCHLOROQUIN/AZITHROMYCIN PROTOCOL ON COVID-19 CASE-FATALITY RATE REDUCTION IN ALGERIA, 3 (2020) 66–72.

[9] N. Kadi, M. Khelfaoui, Population density, a factor in the spread of COVID-19 in Algeria: statistic study., Bull. Natl. Res. Cent. 44 (2020) 138. https://doi.org/10.1186/s42269-020-00393-x.

[10] M. Lounis, D.K. Bagal, Estimation of SIR model's parameters of COVID-19 in Algeria., Bull. Natl. Res. Cent. 44 (2020) 180. https://doi.org/10.1186/s42269-020-00434-5.

[11] S. HananeZermane, DIGITAL LEARNING WITH COVID-19 IN ALGERIA, Journal. 4 (2020) 161–170.

[12] N. Ghounane, H. Rabahi, Moodle in the Algerian EFL Context during Covid-19: Exploring Students' Attitudes and Academic Achievements, Arab World English J. Spec. Issue Commun. Lang. Virtual Spaces. (2023).

[13] A. Chelghoum, H. Chelghoum, The Covid-19 Pandemic and Education: Big Changes ahead for Teaching in Algeria, Altralang J. 2 (2020) 118–132. https://doi.org/10.52919/altralang.v2i02.79.

[14] M. Ciotti, M. Ciccozzi, A. Terrinoni, W.C. Jiang, C. Bin Wang, S. Bernardini, The COVID-19 pandemic, Crit. Rev. Clin. Lab. Sci. (2020) 365–388. https://doi.org/10.1080/10408363.2020.1783198.

[15] Y. Alimohamadi, M. Sepandi, M. Taghdir, H. Hosamirudsari, Determine the most common clinical symptoms in COVID-19 patients: a systematic review and meta-analysis., J. Prev. Med. Hyg. 61 (2020) E304–E312. https://doi.org/10.15167/2421-4248/jpmh2020.61.3.1530.

[16] N. Zhu, D. Zhang, W. Wang, X. Li, B. Yang, J. Song, X. Zhao, B. Huang, W. Shi, R. Lu, P. Niu, F. Zhan, X. Ma, D. Wang, W. Xu, G. Wu, G.F. Gao, W. Tan, A Novel Coronavirus from Patients with Pneumonia in China, 2019, N. Engl. J. Med. 382 (2020) 727–733. https://doi.org/10.1056/NEJMoa2001017.

[17] D. Wang, B. Hu, C. Hu, F. Zhu, X. Liu, J. Zhang, B. Wang, H. Xiang, Z. Cheng, Y. Xiong, Y. Zhao, Y. Li, X. Wang, Z. Peng, Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China., JAMA. 323 (2020) 1061–1069. https://doi.org/10.1001/jama.2020.1585.

[18] R.S. QUAZI, A.A. Joshi, C.B. Hangargekar, R.S. QUAZI, A.A. Joshi, A REVIEW ON COVID-19-A GLOBAL BATTLE BETWEEN LIFE AND DEATH, Int. J. Curr. Pharm. Res. 12 (2020) 19–24. https://doi.org/10.22159/ijcpr.2020v12i4.39084.

[19] N. Chen, M. Zhou, X. Dong, J. Qu, F. Gong, Y. Han, Y. Qiu, J. Wang, Y. Liu, Y. Wei, J. Xia, T. Yu, X. Zhang, L. Zhang, Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study., Lancet (London, England). 395 (2020) 507–513. https://doi.org/10.1016/S0140-6736(20)30211-7.

[20] G. Lippi, M. Plebani, The critical role of laboratory medicine during coronavirus disease 2019 (COVID-19) and other viral outbreaks, Clin. Chem. Lab. Med. 58 (2020). https://doi.org/10.1515/cclm-2020-0240.

[21] C.-Z. Wang, S.-L. Hu, L. Wang, M. Li, H.-T. Li, Early risk factors of the exacerbation of coronavirus disease 2019 pneumonia., J. Med. Virol. 92 (2020) 2593–2599. https://doi.org/10.1002/jmv.26071.

[22] A. Bhargava, E.A. Fukushima, M. Levine, W. Zhao, F. Tanveer, S.M. Szpunar, L. Saravolatz, Predictors for Severe COVID-19 Infection., Clin. Infect. Dis. an Off. Publ. Infect. Dis. Soc. Am. 71 (2020) 1962–1968. https://doi.org/10.1093/cid/ciaa674.

[23] I. Hamming, W. Timens, M.L.C. Bulthuis, A.T. Lely, G.J. Navis, H. van Goor, Tissue distribution of ACE2 protein, the functional receptor for SARS coronavirus. A first step in understanding SARS pathogenesis., J. Pathol. 203 (2004) 631–637. https://doi.org/10.1002/path.1570.

[24] K. Renu, P.L. Prasanna, A. Valsala Gopalakrishnan, Coronaviruses pathogenesis, comorbidities and multi-organ damage - A review., Life Sci. 255 (2020) 117839. https://doi.org/10.1016/j.lfs.2020.117839.

[25] B. Long, W.J. Brady, A. Koyfman, M. Gottlieb, Cardiovascular complications in COVID-19., Am. J. Emerg. Med. 38 (2020) 1504–1507. https://doi.org/10.1016/j.ajem.2020.04.048.

[26] C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, L. Zhang, G. Fan, J. Xu, X. Gu, Z. Cheng, T. Yu, J. Xia, Y. Wei, W. Wu, X. Xie, W. Yin, H. Li, M. Liu, Y. Xiao, H. Gao, L. Guo, J. Xie, G. Wang, R. Jiang, Z. Gao, Q. Jin, J. Wang, B. Cao, Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China., Lancet (London, England). 395 (2020) 497–506. https://doi.org/10.1016/S0140-6736(20)30183-5.

[27] Q. Ruan, K. Yang, W. Wang, L. Jiang, J. Song, Clinical predictors of mortality due to COVID-19 based on an analysis of data of 150 patients from Wuhan, China, Intensive

Care Med. 46 (2020) 846–848. https://doi.org/10.1007/s00134-020-05991-x.

[28] B.M. Henry, S.W. Benoit, M.H.S. de Oliveira, W.C. Hsieh, J. Benoit, R.A. Ballout, M. Plebani, G. Lippi, Laboratory abnormalities in children with mild and severe coronavirus disease 2019 (COVID-19): A pooled analysis and review., Clin. Biochem. 81 (2020) 1–8. https://doi.org/10.1016/j.clinbiochem.2020.05.012.

[29] G. Sanna, G. Serrau, P.P. Bassareo, P. Neroni, V. Fanos, M.A. Marcialis, Children's heart and COVID-19: Up-to-date evidence in the form of a systematic review., Eur. J. Pediatr. 179 (2020) 1079–1087. https://doi.org/10.1007/s00431-020-03699-0.

[30] S.A. Mahmoud, S. Ganesan, E. Ibrahim, B. Thakre, J.G. Teddy, P. Raheja, W.A. Zaher, Evaluation of six different rapid methods for nucleic acid detection of SARS-COV-2 virus, J. Med. Virol. 93 (2021) 5538–5543.

[31] Y. Pan, D. Zhang, P. Yang, L.L.M. Poon, Q. Wang, Viral load of SARS-CoV-2 in clinical samples, Lancet Infect. Dis. 20 (2020) 411–412.

[32] H. Han, Q. Luo, F. Mo, L. Long, W. Zheng, SARS-CoV-2 RNA more readily detected in induced sputum than in throat swabs of convalescent COVID-19 patients, Lancet Infect. Dis. 20 (2020) 655–656.

[33] L. Zou, F. Ruan, M. Huang, L. Liang, H. Huang, Z. Hong, J. Yu, M. Kang, Y. Song, J. Xia, SARS-CoV-2 viral load in upper respiratory specimens of infected patients, N. Engl. J. Med. 382 (2020) 1177–1179.

[34] W. Zhang, R.-H. Du, B. Li, X.-S. Zheng, X.-L. Yang, B. Hu, Y.-Y. Wang, G.-F. Xiao, B. Yan, Z.-L. Shi, Molecular and serological investigation of 2019-nCoV infected patients: implication of multiple shedding routes, Emerg. Microbes Infect. 9 (2020) 386–389.

[35] R. Wolfel, V.M. Corman, W. Guggemos, M. Seilmaier, S. Zange, M.A. Muller, D. Niemeyer, Author Correction: Virological assessment of hospitalized patients with COVID-2019, Nature. 588 (2020) E35–E35.

[36] J.P. Kanne, Chest CT findings in 2019 novel coronavirus (2019-nCoV) infections from Wuhan, China: key points for the radiologist, Radiology. 295 (2020) 16–17.

[37] X. Xie, Z. Zhong, W. Zhao, C. Zheng, F. Wang, J. Liu, Chest CT for typical coronavirus disease 2019 (COVID-19) pneumonia: relationship to negative RT-PCR testing, Radiology. 296 (2020) E41–E45.

[38] A. La Marca, M. Capuzzo, T. Paglia, L. Roli, T. Trenti, S.M. Nelson, Testing for SARS-CoV-2 (COVID-19): a systematic review and clinical guide to molecular and serological in-vitro diagnostic assays, Reprod. Biomed. Online. 41 (2020) 483–499.

[39] S.H. Zhan, S.M. Alamouti, H. Daneshpajouh, B.S. Kwok, M.-H. Lee, J. Khattra, H.J. Houck, K.H. Rand, Target capture sequencing of SARS-CoV-2 genomes using the ONETest Coronaviruses Plus., Diagn. Microbiol. Infect. Dis. 101 (2021) 115508. https://doi.org/10.1016/j.diagmicrobio.2021.115508.

[40] P. De Benedictis, S. Schultz-Cherry, A. Burnham, G. Cattoli, Astrovirus infections in humans and animals–molecular biology, genetic diversity, and interspecies transmissions, Infect. Genet. Evol. 11 (2011) 1529–1544.

[41] K. Dhama, S.K. Patel, K. Sharun, M. Pathak, R. Tiwari, M.I. Yatoo, Y.S. Malik, R. Sah, A.A. Rabaan, P.K. Panwar, K.P. Singh, I. Michalak, W. Chaicumpa, D.F. Martinez-Pulgarin, D.K. Bonilla-Aldana, A.J. Rodriguez-Morales, SARS-CoV-2 jumping the species barrier: Zoonotic lessons from SARS, MERS and recent advances to combat this pandemic virus, Travel Med. Infect. Dis. 37 (2020) 101830. https://doi.org/https://doi.org/10.1016/j.tmaid.2020.101830.

[42] A. Shah, R. Kashyap, P. Tosh, P. Sampathkumar, J.C. O'Horo, Guide to understanding the 2019 novel coronavirus, in: Mayo Clin. Proc., Elsevier, 2020: pp. 646–652.

[43] A. Madani, O. Benkortbi, M. Laidi, In silico prediction of the inhibition of new molecules on SARS-CoV-2 3CL protease by using QSAR: PSOSVR approach, Brazilian J. Chem. Eng. (2023). https://doi.org/10.1007/s43153-023-00332-z.

[44] T.V. Nguyen, Q. Dai Tran, L.T. Phan, L.N. Vu, D.T.T. Truong, H.C. Truong, T.N. Le, L.D.K. Vien, T.V. Nguyen, Q.C. Luong, In the interest of public safety: rapid response to the COVID-19 epidemic in Vietnam, BMJ Glob. Heal. 6 (2021) e004100.

[45] S.C. Pandey, V. Pande, D. Sati, S. Upreti, M. Samant, Vaccination strategies to combat novel corona virus SARS-CoV-2, Life Sci. 256 (2020) 117956. https://doi.org/https://doi.org/10.1016/j.lfs.2020.117956.

[46] S. Mandal, T. Roychowdhury, A. Bhattacharya, Pattern of genomic variation in SARS-CoV-2 (COVID-19) suggests restricted nonrandom changes: Analysis using Shewhart control charts, J. Biosci. 46 (2021) 1–7.

[47] T. Acter, N. Uddin, J. Das, A. Akhter, T.R. Choudhury, S. Kim, Evolution of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) as coronavirus disease 2019 (COVID-19) pandemic: A global health emergency, Sci. Total Environ. 730 (2020) 138996.

[48] R. Lu, X. Zhao, J. Li, P. Niu, B. Yang, H. Wu, W. Wang, H. Song, B. Huang, N. Zhu, Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding, Lancet. 395 (2020) 565–574.

[49] J.F.-W. Chan, K.-H. Kok, Z. Zhu, H. Chu, K.K.-W. To, S. Yuan, K.-Y. Yuen, Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan, Emerg. Microbes Infect. 9 (2020) 221–236.

[50] X. Tang, C. Wu, X. Li, Y. Song, X. Yao, X. Wu, Y. Duan, H. Zhang, Y. Wang, Z. Qian, On the origin and continuing evolution of SARS-CoV-2, Natl. Sci. Rev. 7 (2020) 1012–1023.

[51] J. Damas, G.M. Hughes, K.C. Keough, C.A. Painter, N.S. Persky, M. Corbo, M. Hiller, K.-P. Koepfli, A.R. Pfenning, H. Zhao, Broad host range of SARS-CoV-2 predicted by comparative and structural analysis of ACE2 in vertebrates, Proc. Natl. Acad. Sci. 117 (2020) 22311–22322.

[52] R. Cagliani, D. Forni, M. Clerici, M. Sironi, Computational inference of selection underlying the evolution of the novel coronavirus, severe acute respiratory syndrome coronavirus 2, J. Virol. 94 (2020) 10–1128.

[53] P. Zhou, X. Lou Yang, X.G. Wang, B. Hu, L. Zhang, W. Zhang, H.R. Si, Y. Zhu, B.

Li, C.L. Huang, H.D. Chen, J. Chen, Y. Luo, H. Guo, R. Di Jiang, M.Q. Liu, Y. Chen, X.R. Shen, X. Wang, X.S. Zheng, K. Zhao, Q.J. Chen, F. Deng, L.L. Liu, B. Yan, F.X. Zhan, Y.Y. Wang, G.F. Xiao, Z.L. Shi, A pneumonia outbreak associated with a new coronavirus of probable bat origin, Nature. 579 (2020) 270–273. https://doi.org/10.1038/s41586-020-2012-7.

[54] R.F. Garry, The proximal origin of SARS-CoV-2, Nat. Med. 26 (2020) 450–452.

[55] B. Coutard, C. Valle, X. De Lamballerie, B. Canard, N.G. Seidah, E. Decroly, The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade, Antiviral Res. 176 (2020) 104742.

[56] H. Zhou, X. Chen, T. Hu, J. Li, H. Song, Y. Liu, P. Wang, D. Liu, J. Yang, E.C. Holmes, A novel bat coronavirus closely related to SARS-CoV-2 contains natural insertions at the S1/S2 cleavage site of the spike protein, Curr. Biol. 30 (2020) 2196–2203.

[57] A.G. Wrobel, D.J. Benton, P. Xu, C. Roustan, S.R. Martin, P.B. Rosenthal, J.J. Skehel, S.J. Gamblin, SARS-CoV-2 and bat RaTG13 spike glycoprotein structures inform on virus evolution and furin-cleavage effects, Nat. Struct. Mol. Biol. 27 (2020) 763–767.

[58] Y.C.F. Su, D.E. Anderson, B.E. Young, M. Linster, F. Zhu, J. Jayakumar, Y. Zhuang, S. Kalimuddin, J.G.H. Low, C.W. Tan, Discovery and genomic characterization of a 382-nucleotide deletion in ORF7b and ORF8 during the early evolution of SARS-CoV-2, MBio. 11 (2020) 10–1128.

[59] W.-M. Zhao, S.-H. Song, M.-L. Chen, D. Zou, L.-N. Ma, Y.-K. Ma, R.-J. Li, L.-L. Hao, C.-P. Li, D.-M. Tian, The 2019 novel coronavirus resource., Yi Chuan= Hered. 42 (2020) 212–221.

[60] B. Korber, W.M. Fischer, S. Gnanakaran, H. Yoon, J. Theiler, W. Abfalterer, N. Hengartner, E.E. Giorgi, T. Bhattacharya, B. Foley, Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus, Cell. 182 (2020) 812–827.

[61] P. Forster, L. Forster, C. Renfrew, M. Forster, Phylogenetic network analysis of SARS-CoV-2 genomes, Proc. Natl. Acad. Sci. 117 (2020) 9241–9243.

[62] Q. Han, Q. Lin, Z. Ni, L. You, Uncertainties about the transmission routes of 2019 novel coronavirus, Influenza Other Respi. Viruses. 14 (2020) 470–471. https://doi.org/https://doi.org/10.1111/irv.12735.

[63] N.H.L. Leung, D.K.W. Chu, E.Y.C. Shiu, K.-H. Chan, J.J. McDevitt, B.J.P. Hau, H.-L. Yen, Y. Li, D.K.M. Ip, J.S.M. Peiris, W.-H. Seto, G.M. Leung, D.K. Milton, B.J. Cowling, Respiratory virus shedding in exhaled breath and efficacy of face masks, Nat. Med. 26 (2020) 676–680. https://doi.org/10.1038/s41591-020-0843-2.

[64] M.L. Holshue, C. DeBolt, S. Lindquist, K.H. Lofy, J. Wiesman, H. Bruce, C. Spitters, K. Ericson, S. Wilkerson, A. Tural, G. Diaz, A. Cohn, L. Fox, A. Patel, S.I. Gerber, L. Kim, S. Tong, X. Lu, S. Lindstrom, M.A. Pallansch, W.C. Weldon, H.M. Biggs, T.M. Uyeki, S.K. Pillai, First Case of 2019 Novel Coronavirus in the United States., N. Engl. J. Med. 382 (2020) 929–936. https://doi.org/10.1056/NEJMoa2001191.

[65] A. Núñez-Delgado, What do we know about the SARS-CoV-2 coronavirus in the

environment?, Sci. Total Environ. 727 (2020) 138647. https://doi.org/10.1016/j.scitotenv.2020.138647.

[66] K.K.-W. To, O.T.-Y. Tsang, C.C.-Y. Yip, K.-H. Chan, T.-C. Wu, J.M.-C. Chan, W.-S. Leung, T.S.-H. Chik, C.Y.-C. Choi, D.H. Kandamby, D.C. Lung, A.R. Tam, R.W.-S. Poon, A.Y.-F. Fung, I.F.-N. Hung, V.C.-C. Cheng, J.F.-W. Chan, K.-Y. Yuen, Consistent Detection of 2019 Novel Coronavirus in Saliva., Clin. Infect. Dis. an Off. Publ. Infect. Dis. Soc. Am. 71 (2020) 841–843. https://doi.org/10.1093/cid/ciaa149.

[67] L. Liu, Q. Wei, X. Alvarez, H. Wang, Y. Du, H. Zhu, H. Jiang, J. Zhou, P. Lam, L. Zhang, A. Lackner, C. Qin, Z. Chen, Epithelial cells lining salivary gland ducts are early target cells of severe acute respiratory syndrome coronavirus infection in the upper respiratory tracts of rhesus macaques., J. Virol. 85 (2011) 4025–4030. https://doi.org/10.1128/JVI.02292-10.

[68] V.W.-S. Chan, P.K.-F. Chiu, C.-H. Yee, Y. Yuan, C.-F. Ng, J.Y.-C. Teoh, A systematic review on COVID-19: urological manifestations, viral RNA detection and special considerations in urological conditions., World J. Urol. 39 (2021) 3127–3138. https://doi.org/10.1007/s00345-020-03246-4.

[69] H. Chen, J. Guo, C. Wang, F. Luo, X. Yu, W. Zhang, J. Li, D. Zhao, D. Xu, Q. Gong, J. Liao, H. Yang, W. Hou, Y. Zhang, Clinical characteristics and intrauterine vertical transmission potential of COVID-19 infection in nine pregnant women: a retrospective review of medical records, Lancet. 395 (2020) 809–815. https://doi.org/10.1016/S0140-6736(20)30360-3.

[70] L. Dong, J. Tian, S. He, C. Zhu, J. Wang, C. Liu, J. Yang, Possible Vertical Transmission of SARS-CoV-2 From an Infected Mother to Her Newborn., JAMA. 323 (2020) 1846–1848. https://doi.org/10.1001/jama.2020.4621.

[71] F. Colavita, D. Lapa, F. Carletti, E. Lalle, L. Bordi, P. Marsella, E. Nicastri, N. Bevilacqua, M.L. Giancola, A. Corpolongo, G. Ippolito, M.R. Capobianchi, C. Castilletti, SARS-CoV-2 Isolation From Ocular Secretions of a Patient With COVID-19 in Italy With Prolonged Viral RNA Detection., Ann. Intern. Med. 173 (2020) 242–243. https://doi.org/10.7326/M20-1176.

[72] C.-W. Lu, X.-F. Liu, Z.-F. Jia, 2019-nCoV transmission through the ocular surface must not be ignored., Lancet (London, England). 395 (2020) e39. https://doi.org/10.1016/S0140-6736(20)30313-5.

[73] C.B. Jackson, M. Farzan, B. Chen, H. Choe, Mechanisms of SARS-CoV-2 entry into cells, Nat. Rev. Mol. Cell Biol. 23 (2022) 3–20. https://doi.org/10.1038/s41580-021-00418-x.

[74] S. Belouzard, V.C. Chu, G.R. Whittaker, Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites, Proc. Natl. Acad. Sci. 106 (2009) 5871–5876.

[75] L. Zhang, C.B. Jackson, H. Mou, A. Ojha, H. Peng, B.D. Quinlan, E.S. Rangarajan, A. Pan, A. Vanderheiden, M.S. Suthar, SARS-CoV-2 spike-protein D614G mutation increases virion spike density and infectivity, Nat. Commun. 11 (2020) 6013.

[76] J. Zhang, Y. Cai, T. Xiao, J. Lu, H. Peng, S.M. Sterling, R.M. Walsh Jr, S. Rits-

Volloch, H. Zhu, A.N. Woosley, Structural impact on SARS-CoV-2 spike protein by D614G substitution, Science (80-.). 372 (2021) 525–530.

[77] T.P. Peacock, D.H. Goldhill, J. Zhou, L. Baillon, R. Frise, O.C. Swann, R. Kugathasan, R. Penn, J.C. Brown, R.Y. Sanchez-David, The furin cleavage site in the SARS-CoV-2 spike protein is required for transmission in ferrets, Nat. Microbiol. 6 (2021) 899–909.

[78] S. Matsuyama, N. Nagata, K. Shirato, M. Kawase, M. Takeda, F. Taguchi, Efficient activation of the severe acute respiratory syndrome coronavirus spike protein by the transmembrane protease TMPRSS2, J. Virol. 84 (2010) 12658–12664.

[79] I. Glowacka, S. Bertram, M.A. Müller, P. Allen, E. Soilleux, S. Pfefferle, I. Steffen, T.S. Tsegaye, Y. He, K. Gnirss, Evidence that TMPRSS2 activates the severe acute respiratory syndrome coronavirus spike protein for membrane fusion and reduces viral control by the humoral immune response, J. Virol. 85 (2011) 4122–4134.

[80] A.C. Walls, M.A. Tortorici, J. Snijder, X. Xiong, B.-J. Bosch, F.A. Rey, D. Veesler, Tectonic conformational changes of a coronavirus spike glycoprotein promote membrane fusion, Proc. Natl. Acad. Sci. 114 (2017) 11157–11162.

[81] X. Fan, D. Cao, L. Kong, X. Zhang, Cryo-EM analysis of the post-fusion structure of the SARS-CoV spike glycoprotein, Nat. Commun. 11 (2020) 3618.

[82] M.M. Shaik, H. Peng, J. Lu, S. Rits-Volloch, C. Xu, M. Liao, B. Chen, Structural basis of coreceptor recognition by HIV-1 envelope spike, Nature. 565 (2019) 318–323.

[83] P. V'kovski, A. Kratzel, S. Steiner, H. Stalder, V. Thiel, Coronavirus biology and replication: implications for SARS-CoV-2, Nat. Rev. Microbiol. 19 (2021) 155–170.

[84] J. Emrani, M. Ahmed, L. Jeffers-Francis, J.C. Teleha, N. Mowa, R.H. Newman, M.D. Thomas, SARS-COV-2, infection, transmission, transcription, translation, proteins, and treatment: A review., Int. J. Biol. Macromol. 193 (2021) 1249–1273. https://doi.org/10.1016/j.ijbiomac.2021.10.172.

[85] Y. Finkel, O. Mizrahi, A. Nachshon, S. Weingarten-Gabbay, D. Morgenstern, Y. Yahalom-Ronen, H. Tamir, H. Achdout, D. Stein, O. Israeli, The coding capacity of SARS-CoV-2, Nature. 589 (2021) 125–130.

[86] N.S. Ogando, J.C. Zevenhoven-Dobbe, Y. van der Meer, P.J. Bredenbeek, C.C. Posthuma, E.J. Snijder, The enzymatic activity of the nsp14 exoribonuclease is critical for replication of MERS-CoV and SARS-CoV-2, J. Virol. 94 (2020) 10–1128.

[87] S. Satarker, M. Nampoothiri, Structural proteins in severe acute respiratory syndrome coronavirus-2, Arch. Med. Res. 51 (2020) 482–491.

[88] C.J. Budzilowicz, S.P. Wilczynski, S.R. Weiss, Three intergenic regions of coronavirus mouse hepatitis virus strain A59 genome RNA contain a common nucleotide sequence that is homologous to the 3'end of the viral mRNA leader sequence, J. Virol. 53 (1985) 834–840.

[89] S.G. Sawicki, D.L. Sawicki, Coronavirus minus-strand RNA synthesis and effect of cycloheximide on coronavirus RNA synthesis, J. Virol. 57 (1986) 328–334.

[90] T. Viswanathan, S. Arya, S.-H. Chan, S. Qi, N. Dai, A. Misra, J.-G. Park, F. Oladunni,

D. Kovalskyy, R.A. Hromas, Structural basis of RNA cap modification by SARS-CoV-2, Nat. Commun. 11 (2020) 3718.

[91] A.M. Collier, O.L. Lyytinen, Y.R. Guo, Y. Toh, M.M. Poranen, Y.J. Tao, Initiation of RNA polymerization and polymerase encapsidation by a small dsRNA virus, PLoS Pathog. 12 (2016) e1005523.

[92] I. Alam, A.A. Kamau, M. Kulmanov, Ł. Jaremko, S.T. Arold, A. Pain, T. Gojobori, C.M. Duarte, Functional pangenome analysis shows key features of E protein are preserved in SARS and SARS-CoV-2, Front. Cell. Infect. Microbiol. 10 (2020) 405.

[93] J.L. Nieva, V. Madan, L. Carrasco, Viroporins: structure and biological functions, Nat. Rev. Microbiol. 10 (2012) 563–574.

[94] L. Wilson, C. Mckinlay, P. Gage, G. Ewart, SARS coronavirus E protein forms cation-selective ion channels, Virology. 330 (2004) 322–331.

[95] T. Pham, J.L. Perry, T.L. Dosey, A.H. Delcour, J.M. Hyser, The rotavirus NSP4 viroporin domain is a calcium-conducting ion channel, Sci. Rep. 7 (2017) 43487.

[96] R. Zhang, K. Wang, W. Lv, W. Yu, S. Xie, K. Xu, W. Schwarz, S. Xiong, B. Sun, The ORF4a protein of human coronavirus 229E functions as a viroporin that regulates viral production, Biochim. Biophys. Acta (BBA)-Biomembranes. 1838 (2014) 1088–1095.

[97] Q. Wu, Y. Zhang, H. Lü, J. Wang, X. He, Y. Liu, C. Ye, W. Lin, J. Hu, J. Ji, The E protein is a multifunctional membrane protein of SARS-CoV, Genomics. Proteomics Bioinformatics. 1 (2003) 131–144.

[98] Y. Liao, J.P. Tam, D.X. Liu, Viroporin activity of SARS-CoV E protein, in: Nidoviruses Towar. Control SARS Other Nidovirus Dis., Springer, 2006: pp. 199–202.

[99] G. Duart, M.J. García-Murria, B. Grau, J.M. Acosta-Cáceres, L. Martínez-Gil, I. Mingarro, SARS-CoV-2 envelope protein topology in eukaryotic membranes, Open Biol. 10 (2020) 200209.

[100] T.S. Fung, D.X. Liu, Coronavirus infection, ER stress, apoptosis and innate immunity, Front. Microbiol. 5 (2014) 296.

[101] M. Bianchi, D. Benvenuto, M. Giovanetti, S. Angeletti, M. Ciccozzi, S. Pascarella, Sars-CoV-2 envelope and membrane proteins: structural differences linked to virus characteristics?, Biomed Res. Int. 2020 (2020).

[102] A.L. Arndt, B.J. Larson, B.G. Hogue, A conserved domain in the coronavirus membrane protein tail is important for virus assembly, J. Virol. 84 (2010) 11418–11428.

[103] L. Zhang, D. Lin, X. Sun, U. Curth, C. Drosten, L. Sauerhering, S. Becker, K. Rox, R. Hilgenfeld, Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors, Science (80-.). 368 (2020) 409–412.

[104] J. Lan, J. Ge, J. Yu, S. Shan, H. Zhou, S. Fan, Q. Zhang, X. Shi, Q. Wang, L. Zhang, Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor, Nature. 581 (2020) 215–220.

[105] R. McBride, M. Van Zyl, B.C. Fielding, The coronavirus nucleocapsid is a multifunctional protein, Viruses. 6 (2014) 2991–3018.

[106] M. Surjit, S.K. Lal, The SARS-CoV nucleocapsid protein: a protein with multifarious activities, Infect. Genet. Evol. 8 (2008) 397–405.

[107] S. Kang, M. Yang, Z. Hong, L. Zhang, Z. Huang, X. Chen, S. He, Z. Zhou, Z. Zhou, Q. Chen, Crystal structure of SARS-CoV-2 nucleocapsid protein RNA binding domain reveals potential unique drug targeting sites, Acta Pharm. Sin. B. 10 (2020) 1228–1238.

[108] A. Vandelli, M. Monti, E. Milanetti, A. Armaos, J. Rupert, E. Zacco, E. Bechara, R. Delli Ponti, G.G. Tartaglia, Structural analysis of SARS-CoV-2 genome and predictions of the human interactome, Nucleic Acids Res. 48 (2020) 11270–11283.

[109] K. Schubert, E.D. Karousis, A. Jomaa, A. Scaiola, B. Echeverria, L.-A. Gurzeler, M. Leibundgut, V. Thiel, O. Mühlemann, N. Ban, SARS-CoV-2 Nsp1 binds the ribosomal mRNA channel to inhibit translation, Nat. Struct. Mol. Biol. 27 (2020) 959–966.

[110] M. Thoms, R. Buschauer, M. Ameismeier, L. Koepke, T. Denk, M. Hirschenberger, H. Kratzat, M. Hayn, T. Mackens-Kiani, J. Cheng, Structural basis for translational shutdown and immune evasion by the Nsp1 protein of SARS-CoV-2, Science (80-.). 369 (2020) 1249–1255.

[111] C.T. Cornillez-Ty, L. Liao, J.R. Yates III, P. Kuhn, M.J. Buchmeier, Severe acute respiratory syndrome coronavirus nonstructural protein 2 interacts with a host protein complex involved in mitochondrial biogenesis and intracellular signaling, J. Virol. 83 (2009) 10314–10318.

[112] J. Lei, Y. Kusov, R. Hilgenfeld, Nsp3 of coronaviruses: Structures and functions of a large multi-domain protein, Antiviral Res. 149 (2018) 58–74.

[113] J.-M. Claverie, A putative role of de-mono-ADP-ribosylation of STAT1 by the SARS-CoV-2 Nsp3 protein in the cytokine storm syndrome of COVID-19, Viruses. 12 (2020) 646.

[114] G. Li, E. De Clercq, Therapeutic options for the 2019 novel coronavirus (2019-nCoV), Nat. Rev. Drug Discov. 19 (2020) 149–150.

[115] K.C. Lehmann, A. Gulyaeva, J.C. Zevenhoven-Dobbe, G.M.C. Janssen, M. Ruben, H.S. Overkleeft, P.A. Van Veelen, D. V Samborskiy, A.A. Kravchenko, A.M. Leontovich, Discovery of an essential nucleotidylating activity associated with a newly delineated conserved domain in the RNA polymerase-containing protein of all nidoviruses, Nucleic Acids Res. 43 (2015) 8416–8434.

[116] R.N. Kirchdoerfer, A.B. Ward, Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors, Nat. Commun. 10 (2019) 2342.

[117] Y. Ma, L. Wu, N. Shaw, Y. Gao, J. Wang, Y. Sun, Z. Lou, L. Yan, R. Zhang, Z. Rao, Structural basis and functional analysis of the SARS coronavirus nsp14–nsp10 complex, Proc. Natl. Acad. Sci. 112 (2015) 9436–9441.

[118] Y. Kim, R. Jedrzejczak, N.I. Maltseva, M. Wilamowski, M. Endres, A. Godzik, K. Michalska, A. Joachimiak, Crystal structure of Nsp15 endoribonuclease NendoU from SARS-CoV-2, Protein Sci. 29 (2020) 1596–1605.

[119] Y. Dongliang, S.S. Ejaz, M. Aziz, A. Saeed, S.S. Ejaz, M. Bilal, M. Kashif, S.S. Ejaz, Benzene-1,3-diol derivatives as the inhibitors of butyrylcholinesterase: An emergent target of Alzheimer's disease, J. Serbian Chem. Soc. 87 (2022) 293–306. https://doi.org/10.2298/jsc210416073d.

[120] P. De, S. Bhayye, V. Kumar, K. Roy, In silico modeling for quick prediction of inhibitory activity against 3CLpro enzyme in SARS CoV diseases, J. Biomol. Struct. Dyn. 0 (2020) 1–27. https://doi.org/10.1080/07391102.2020.1821779.

[121] F. Yan, F. Gao, An overview of potential inhibitors targeting non-structural proteins 3 (PLpro and Mac1) and 5 (3CLpro/Mpro) of SARS-CoV-2, Comput. Struct. Biotechnol. J. 19 (2021) 4868–4883. https://doi.org/10.1016/j.csbj.2021.08.036.

[122] K. Narayanan, C. Huang, S. Makino, SARS coronavirus accessory proteins, Virus Res. 133 (2008) 113–121.

[123] A. Wu, Y. Peng, B. Huang, X. Ding, X. Wang, P. Niu, J. Meng, Z. Zhu, Z. Zhang, J. Wang, Genome composition and divergence of the novel coronavirus (2019-nCoV) originating in China, Cell Host Microbe. 27 (2020) 325–328.

[124] C.A.M. De Haan, L. Kuo, P.S. Masters, H. Vennema, P.J.M. Rottier, Coronavirus particle assembly: primary structure requirements of the membrane protein, J. Virol. 72 (1998) 6838–6850.

[125] M.A. Shereen, S. Khan, A. Kazmi, N. Bashir, R. Siddique, COVID-19 infection: Emergence, transmission, and characteristics of human coronaviruses, J. Adv. Res. 24 (2020) 91–98.

[126] J. Yang, R. Yan, A. Roy, D. Xu, J. Poisson, Y. Zhang, The I-TASSER Suite: protein structure and function prediction, Nat. Methods. 12 (2015) 7–8.

[127] A. Bartesaghi, D. Matthies, S. Banerjee, A. Merk, S. Subramaniam, Structure of β-galactosidase at 3.2-Å resolution obtained by cryo-electron microscopy, Proc. Natl. Acad. Sci. 111 (2014) 11709–11714.

[128] B. Cosar, Z.Y. Karagulleoglu, S. Unal, A.T. Ince, D.B. Uncuoglu, G. Tuncer, B.R. Kilinc, Y.E. Ozkan, H.C. Ozkoc, I.N. Demir, A. Eker, F. Karagoz, S.Y. Simsek, B. Yasar, M. Pala, A. Demir, I.N. Atak, A.H. Mendi, V.U. Bengi, G. Cengiz Seval, E. Gunes Altuntas, P. Kilic, D. Demir-Dora, SARS-CoV-2 Mutations and their Viral Variants, Cytokine Growth Factor Rev. 63 (2022) 10–22. https://doi.org/https://doi.org/10.1016/j.cytogfr.2021.06.001.

[129] J. Chen, R. Wang, M. Wang, G.-W. Wei, Mutations strengthened SARS-CoV-2 infectivity, J. Mol. Biol. 432 (2020) 5212–5226.

[130] A.M. Syed, T.Y. Taha, T. Tabata, I.P. Chen, A. Ciling, M.M. Khalid, B. Sreekumar, P.-Y. Chen, J.M. Hayashi, K.M. Soczek, Rapid assessment of SARS-CoV-2–evolved variants using virus-like particles, Science (80-.). 374 (2021) 1626–1632.

[131] H. Wu, N. Xing, K. Meng, B. Fu, W. Xue, P. Dong, W. Tang, Y. Xiao, G. Liu, H. Luo, Nucleocapsid mutations R203K/G204R increase the infectivity, fitness, and virulence of SARS-CoV-2, Cell Host Microbe. 29 (2021) 1788–1801.

[132] B.A. Johnson, Y. Zhou, K.G. Lokugamage, M.N. Vu, N. Bopp, P.A. Crocquet-Valdes,

B. Kalveram, C. Schindewolf, Y. Liu, D. Scharton, Nucleocapsid mutations in SARS-CoV-2 augment replication and pathogenesis, PLoS Pathog. 18 (2022) e1010627.

[133] A.M. Carabelli, T.P. Peacock, L.G. Thorne, W.T. Harvey, J. Hughes, T.I. de Silva, S.J. Peacock, W.S. Barclay, T.I. de Silva, G.J. Towers, D.L. Robertson, C.-19 G.U.K. Consortium, SARS-CoV-2 variant biology: immune escape, transmission and fitness, Nat. Rev. Microbiol. 21 (2023) 162–177. https://doi.org/10.1038/s41579-022-00841-7.

[134] H. V Mears, G.R. Young, T. Sanderson, R. Harvey, M. Crawford, D.M. Snell, A.S. Fowler, S. Hussain, J. Nicod, T.P. Peacock, E. Emmott, K. Finsterbusch, J. Luptak, E. Wall, B. Williams, S. Gandhi, C. Swanton, D.L. V Bauer, Emergence of new subgenomic mRNAs in SARS-CoV-2, BioRxiv. (2022) 2022.04.20.488895. https://doi.org/10.1101/2022.04.20.488895.

[135] S. Leary, S. Gaudieri, M.D. Parker, A. Chopra, I. James, S. Pakala, E. Alves, M. John, B.B. Lindsey, A.J. Keeley, Generation of a novel SARS-CoV-2 sub-genomic RNA due to the R203K/G204R variant in nucleocapsid: homologous recombination has potential to change SARS-CoV-2 at both protein and RNA level, Pathog. Immun. 6 (2021) 27.

[136] A.M. Syed, A. Ciling, T.Y. Taha, I.P. Chen, M.M. Khalid, B. Sreekumar, P.-Y. Chen, G.R. Kumar, R. Suryawanshi, I. Silva, Omicron mutations enhance infectivity and reduce antibody neutralization of SARS-CoV-2 virus-like particles, Proc. Natl. Acad. Sci. 119 (2022) e2200592119.

[137] D. Schoeman, B.C. Fielding, Coronavirus envelope protein: current knowledge, Virol. J. 16 (2019) 1–22.

[138] B. Xia, Y. wang, X. pan, X. Cheng, H. Ji, X. Zuo, J. Li, Z. Gao, Why SARS-CoV-2 Omicron variant is milder? A single high-frequency mutation of structural envelope protein matters, BioRxiv. (2022) 2022.02.01.478647. https://doi.org/10.1101/2022.02.01.478647.

[139] S. Ricciardi, A.M. Guarino, L. Giaquinto, E. V Polishchuk, M. Santoro, G. Di Tullio, C. Wilson, F. Panariello, V.C. Soares, S.S.G. Dias, The role of NSP6 in the biogenesis of the SARS-CoV-2 replication organelle, Nature. 606 (2022) 761–768.

[140] X. Shen, H. Tang, C. McDanal, K. Wagh, W. Fischer, J. Theiler, H. Yoon, D. Li, B.F. Haynes, K.O. Sanders, SARS-CoV-2 variant B. 1.1. 7 is susceptible to neutralizing antibodies elicited by ancestral spike vaccines, Cell Host Microbe. 29 (2021) 529–539.

[141] N.G. Davies, S. Abbott, R.C. Barnard, C.I. Jarvis, A.J. Kucharski, J.D. Munday, C.A.B. Pearson, T.W. Russell, D.C. Tully, A.D. Washburne, Estimated transmissibility and impact of SARS-CoV-2 lineage B. 1.1. 7 in England, Science (80-.). 372 (2021) eabg3055.

[142] A. Rambaut, E.C. Holmes, Á. O'Toole, V. Hill, J.T. McCrone, C. Ruis, L. du Plessis, O.G. Pybus, A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology, Nat. Microbiol. 5 (2020) 1403–1407.

[143] F. Konings, M.D. Perkins, J.H. Kuhn, M.J. Pallen, E.J. Alm, B.N. Archer, A. Barakat, T. Bedford, J.N. Bhiman, L. Caly, SARS-CoV-2 Variants of Interest and Concern naming scheme conducive for global discourse, Nat. Microbiol. 6 (2021) 821–823.

[144] N.R. Faria, T.A. Mellan, C. Whittaker, I.M. Claro, D. da S. Candido, S. Mishra,

M.A.E. Crispim, F.C.S. Sales, I. Hawryluk, J.T. McCrone, Genomics and epidemiology of the P. 1 SARS-CoV-2 lineage in Manaus, Brazil, Science (80-.). 372 (2021) 815–821.

[145] P. V Markov, M. Ghafari, M. Beer, K. Lythgoe, P. Simmonds, N.I. Stilianakis, A. Katzourakis, The evolution of SARS-CoV-2, Nat. Rev. Microbiol. 21 (2023) 361–379. https://doi.org/10.1038/s41579-023-00878-2.

[146] M.S. Dhar, R. Marwal, R. Vs, K. Ponnusamy, B. Jolly, R.C. Bhoyar, V. Sardana, S. Naushin, M. Rophina, T.A. Mellan, Genomic characterization and epidemiology of an emerging SARS-CoV-2 variant in Delhi, India, Science (80-.). 374 (2021) 995–999.

[147] F. Campbell, B. Archer, H. Laurenson-Schafer, Y. Jinnai, F. Konings, N. Batra, B. Pavlin, K. Vandemaele, M.D. Van Kerkhove, T. Jombart, Increased transmissibility and global spread of SARS-CoV-2 variants of concern as at June 2021, Eurosurveillance. 26 (2021) 2100509.

[148] R. Viana, S. Moyo, D.G. Amoako, H. Tegally, C. Scheepers, C.L. Althaus, U.J. Anyaneji, P.A. Bester, M.F. Boni, M. Chand, Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in southern Africa, Nature. 603 (2022) 679–686.

[149] K. Tao, P.L. Tzou, J. Nouhin, R.K. Gupta, T. de Oliveira, S.L. Kosakovsky Pond, D. Fera, R.W. Shafer, The biological and clinical significance of emerging SARS-CoV-2 variants, Nat. Rev. Genet. 22 (2021) 757–773.

[150] S.W. Attwood, S.C. Hill, D.M. Aanensen, T.R. Connor, O.G. Pybus, Phylogenetic and phylodynamic approaches to understanding and combating the early SARS-CoV-2 pandemic, Nat. Rev. Genet. 23 (2022) 547–562.

[151] W.T. Harvey, A.M. Carabelli, B. Jackson, R.K. Gupta, E.C. Thomson, E.M. Harrison, C. Ludden, R. Reeve, A. Rambaut, C.-19 G.U.K. (COG-U. Consortium, SARS-CoV-2 variants, spike mutations and immune escape, Nat. Rev. Microbiol. 19 (2021) 409–424.

[152] F. Obermeyer, M. Jankowiak, N. Barkas, S.F. Schaffner, J.D. Pyle, L. Yurkovetskiy, M. Bosso, D.J. Park, M. Babadi, B.L. MacInnis, Analysis of 6.4 million SARS-CoV-2 genomes identifies mutations associated with fitness, Science (80-.). 376 (2022) 1327–1332.

[153] H. Tegally, E. Wilkinson, M. Giovanetti, A. Iranzadeh, V. Fonseca, J. Giandhari, D. Doolabh, S. Pillay, E.J. San, N. Msomi, Detection of a SARS-CoV-2 variant of concern in South Africa, Nature. 592 (2021) 438–443.

[154] T. Gräf, G. Bello, T.M.M. Venas, E.C. Pereira, A.C.D. Paixão, L.R. Appolinario, R.S. Lopes, A.C.D.F. Mendonça, A.S.B. da Rocha, F.C. Motta, Identification of a novel SARS-CoV-2 P. 1 sub-lineage in Brazil provides new insights about the mechanisms of emergence of variants of concern, Virus Evol. 7 (2021) veab091.

[155] V. Hill, L. Du Plessis, T.P. Peacock, D. Aggarwal, R. Colquhoun, A.M. Carabelli, N. Ellaby, E. Gallagher, N. Groves, B. Jackson, The origins and molecular evolution of SARS-CoV-2 lineage B. 1.1. 7 in the UK, Virus Evol. 8 (2022) veac080.

[156] J.T. McCrone, V. Hill, S. Bajaj, R.E. Pena, B.C. Lambert, R. Inward, S. Bhatt, E. Volz, C. Ruis, S. Dellicour, Context-specific emergence and growth of the SARS-CoV-2 Delta variant, Nature. 610 (2022) 154–160.

[157] H. Tegally, M. Moir, J. Everatt, M. Giovanetti, C. Scheepers, E. Wilkinson, K. Subramoney, Z. Makatini, S. Moyo, D.G. Amoako, Emergence of SARS-CoV-2 omicron lineages BA. 4 and BA. 5 in South Africa, Nat. Med. 28 (2022) 1785–1790.

[158] S. Zeghbib, B.A. Somogyi, B. Zana, G. Kemenesi, R. Herczeg, F. Derrar, F. Jakab, The Algerian Chapter of SARS-CoV-2 Pandemic: An Evolutionary, Genetic, and Epidemiological Prospect., Viruses. 13 (2021). https://doi.org/10.3390/v13081525.

[159] T. Menasria, M. Aguilera, Genomic Diversity of SARS-CoV-2 in Algeria and North African Countries: What We Know So Far and What We Expect?, Microorganisms. 10 (2022). https://doi.org/10.3390/microorganisms10020467.

[160] M.-K. Looi, Covid-19: Scientists sound alarm over new BA.2.86 "Pirola" variant, BMJ. 382 (2023) p1964. https://doi.org/10.1136/bmj.p1964.

[161] Callaway E, Why a highly mutated coronavirus variant has scientists on alert, Nature. (2023). https://www.nature.com/articles/d41586-023-02656-9 (accessed August 29, 2023).

[162] N.C. Kyriakidis, A. López-Cortés, E.V. González, A.B. Grimaldos, E.O. Prado, SARS-CoV-2 vaccines strategies: a comprehensive review of phase 3 candidates, Npj Vaccines. 6 (2021) 28. https://doi.org/10.1038/s41541-021-00292-w.

[163] R. Patel, M. Kaki, V.S. Potluri, P. Kahar, D. Khanna, A comprehensive review of SARS-CoV-2 vaccines: Pfizer, Moderna & Johnson & Johnson, Hum. Vaccin. Immunother. 18 (2022) 2002083. https://doi.org/10.1080/21645515.2021.2002083.

[164] P. Tang, M.R. Hasan, H. Chemaitelly, H.M. Yassine, F.M. Benslimane, H.A. Al Khatib, S. AlMukdad, P. Coyle, H.H. Ayoub, Z. Al Kanaani, E. Al Kuwari, A. Jeremijenko, A.H. Kaleeckal, A.N. Latif, R.M. Shaik, H.F. Abdul Rahim, G.K. Nasrallah, M.G. Al Kuwari, H.E. Al Romaihi, A.A. Butt, M.H. Al-Thani, A. Al Khal, R. Bertollini, L.J. Abu-Raddad, BNT162b2 and mRNA-1273 COVID-19 vaccine effectiveness against the SARS-CoV-2 Delta variant in Qatar, Nat. Med. 27 (2021) 2136–2143. https://doi.org/10.1038/s41591-021-01583-4.

[165] V.E. Hillary, S.A. Ceasar, An update on COVID-19: SARS-CoV-2 variants, antiviral drugs, and vaccines, Heliyon. 9 (2023). https://doi.org/10.1016/j.heliyon.2023.e13952.

[166] Y. Unoh, S. Uehara, K. Nakahara, H. Nobori, Y. Yamatsu, S. Yamamoto, Y. Maruyama, Y. Taoda, K. Kasamatsu, T. Suto, K. Kouki, A. Nakahashi, S. Kawashima, T. Sanaki, S. Toba, K. Uemura, T. Mizutare, S. Ando, M. Sasaki, Y. Orba, H. Sawa, A. Sato, T. Sato, T. Kato, Y. Tachibana, Discovery of S-217622, a Noncovalent Oral SARS-CoV-2 3CL Protease Inhibitor Clinical Candidate for Treating COVID-19, J. Med. Chem. 65 (2022) 6499–6512. https://doi.org/10.1021/acs.jmedchem.2c00117.

[167] H. Fan, F. Lou, J. Fan, M. Li, Y. Tong, The emergence of powerful oral anti-COVID-19 drugs in the post-vaccine era, The Lancet Microbe. 3 (2022) e91. https://doi.org/10.1016/S2666-5247(21)00278-0.

[168] A. Shannon, V. Fattorini, B. Sama, B. Selisko, M. Feracci, C. Falcou, P. Gauffre, P. El Kazzi, A. Delpal, E. Decroly, K. Alvarez, C. Eydoux, J.-C. Guillemot, A. Moussa, S.S. Good, P. La Colla, K. Lin, J.-P. Sommadossi, Y. Zhu, X. Yan, H. Shi, F. Ferron, B. Canard, A dual mechanism of action of AT-527 against SARS-CoV-2 polymerase, Nat.

Commun. 13 (2022) 621. https://doi.org/10.1038/s41467-022-28113-1.

[169] Z. Khan, Y. Karataş, A.F. Ceylan, H. Rahman, COVID-19 and therapeutic drugs repurposing in hand: the need for collaborative efforts, Le Pharm. Hosp. Clin. 56 (2021) 3–11.

[170] T.L. Doan, M. Pollastri, M.A. Walters, G.I. Georg, The future of drug repositioning: old drugs, new opportunities, in: Annu. Rep. Med. Chem., Elsevier, 2011: pp. 385–401.

[171] Z. Khan, Y. Karataş, H. Rahman, Anti COVID-19 drugs: need for more clinical evidence and global action, Adv. Ther. 37 (2020) 2575–2579.

[172] W. Zhu, C.Z. Chen, K. Gorshkov, M. Xu, D.C. Lo, W. Zheng, RNA-dependent RNA polymerase as a target for COVID-19 drug discovery, SLAS Discov. Adv. Sci. Drug Discov. 25 (2020) 1141–1151.

[173] D. Siegel, H.C. Hui, E. Doerffler, M.O. Clarke, K. Chun, L. Zhang, S. Neville, E. Carra, W. Lew, B. Ross, Discovery and Synthesis of a Phosphoramidate Prodrug of a Pyrrolo [2, 1-f][triazin-4-amino] Adenine C-Nucleoside (GS-5734) for the Treatment of Ebola and Emerging Viruses, (2017).

[174] S. Mulangu, L.E. Dodd, R.T. Davey Jr, O. Tshiani Mbaya, M. Proschan, D. Mukadi, M. Lusakibanza Manzo, D. Nzolo, A. Tshomba Oloma, A. Ibanda, A randomized, controlled trial of Ebola virus disease therapeutics, N. Engl. J. Med. 381 (2019) 2293–2303.

[175] T.P. Sheahan, A.C. Sims, S.R. Leist, A. Schäfer, J. Won, A.J. Brown, S.A. Montgomery, A. Hogg, D. Babusis, M.O. Clarke, Comparative therapeutic efficacy of remdesivir and combination lopinavir, ritonavir, and interferon beta against MERS-CoV, Nat. Commun. 11 (2020) 222.

[176] T.P. Sheahan, A.C. Sims, R.L. Graham, V.D. Menachery, L.E. Gralinski, J.B. Case, S.R. Leist, K. Pyrc, J.Y. Feng, I. Trantcheva, Broad-spectrum antiviral GS-5734 inhibits both epidemic and zoonotic coronaviruses, Sci. Transl. Med. 9 (2017) eaal3653.

[177] Z.D. Kifle, A.G. Ayele, E.F. Enyew, Drug Repurposing Approach, Potential Drugs, and Novel Drug Targets for COVID-19 Treatment, J. Environ. Public Health. 2021 (2021) 6631721. https://doi.org/10.1155/2021/6631721.

[178] C. Scavone, S. Brusco, M. Bertini, L. Sportiello, C. Rafaniello, A. Zoccoli, L. Berrino, G. Racagni, F. Rossi, A. Capuano, Current pharmacological treatments for COVID-19: What's next?, Br. J. Pharmacol. 177 (2020) 4813–4824.

[179] D. Wichmann, J.-P. Sperhake, M. Lütgehetmann, S. Steurer, C. Edler, A. Heinemann, F. Heinrich, H. Mushumba, I. Kniep, A.S. Schröder, Autopsy findings and venous thromboembolism in patients with COVID-19: a prospective cohort study, Ann. Intern. Med. 173 (2020) 268–277.

[180] K. Sharun, K. Dhama, S.K. Patel, M. Pathak, R. Tiwari, B.R. Singh, R. Sah, D.K. Bonilla-Aldana, A.J. Rodriguez-Morales, H. Leblebicioglu, Ivermectin, a new candidate therapeutic against SARS-CoV-2/COVID-19, Ann. Clin. Microbiol. Antimicrob. 19 (2020) 1–5.

[181] P. Xu, J. Huang, Z. Fan, W. Huang, M. Qi, X. Lin, W. Song, L. Yi, Arbidol/IFN-α2b therapy for patients with corona virus disease 2019: a retrospective multicenter cohort study, Microbes Infect. 22 (2020) 200–205.

[182] J.S. Lee, E.-C. Shin, The type I interferon response in COVID-19: implications for treatment, Nat. Rev. Immunol. 20 (2020) 585–586.

[183] S.S. Unhale, Q.B. Ansar, S. Sanap, S. Thakhre, S. Wadatkar, R. Bairagi, S. Sagrule, K.R. Biyani, A review on corona virus (COVID-19), World J. Pharm. Life Sci. 6 (2020) 109–115.

[184] M.L. Agostini, E.L. Andres, A.C. Sims, R.L. Graham, T.P. Sheahan, X. Lu, E.C. Smith, J.B. Case, J.Y. Feng, R. Jordan, Coronavirus susceptibility to the antiviral remdesivir (GS-5734) is mediated by the viral polymerase and the proofreading exoribonuclease, MBio. 9 (2018) 10–1128.

[185] J. He, L. Hu, X. Huang, C. Wang, Z. Zhang, Y. Wang, D. Zhang, W. Ye, Potential of coronavirus 3C-like protease inhibitors for the development of new anti-SARS-CoV-2 drugs: Insights from structures of protease and inhibitors, Int. J. Antimicrob. Agents. 56 (2020) 106055.

[186] K.M. Wagstaff, H. Sivakumaran, S.M. Heaton, D. Harrich, D.A. Jans, Ivermectin is a specific inhibitor of importin α/β-mediated nuclear import able to inhibit replication of HIV-1 and dengue virus, Biochem. J. 443 (2012) 851–856.

[187] R.J. Reiter, P. Abreu-Gonzalez, P.E. Marik, A. Dominguez-Rodriguez, Therapeutic algorithm for use of melatonin in patients with COVID-19, Front. Med. (2020) 226.

[188] J. Zhang, X. Ma, F. Yu, J. Liu, F. Zou, T. Pan, H. Zhang, Teicoplanin potently blocks the cell entry of 2019-nCoV, Biorxiv. 2020 (2020) 2–5.

[189] N. Mondaini, Phosphodiesterase type 5 inhibitors and COVID-19: are they useful in disease management?, World J. Mens. Health. 38 (2020) 254.

[190] Z. Shi, C.A. Puyo, N-acetylcysteine to combat COVID-19: an evidence review, Ther. Clin. Risk Manag. (2020) 1047–1055.

[191] J.M. Lucas, C. Heinlein, T. Kim, S.A. Hernandez, M.S. Malik, L.D. True, C. Morrissey, E. Corey, B. Montgomery, E. Mostaghel, The androgen-regulated protease TMPRSS2 activates a proteolytic cascade involving components of the tumor microenvironment and promotes prostate cancer metastasis, Cancer Discov. 4 (2014) 1310–1325.

[192] W. Zhang, Y. Zhao, F. Zhang, Q. Wang, T. Li, Z. Liu, J. Wang, Y. Qin, X. Zhang, X. Yan, The use of anti-inflammatory drugs in the treatment of people with severe coronavirus disease 2019 (COVID-19): The Perspectives of clinical immunologists from China, Clin. Immunol. 214 (2020) 108393.

[193] C. Otuechere, Repurposing of chloroquine and its derivative, hydroxychloroquine for COVID-19: implications for people living with HIV in Africa, (2020).

[194] M.R. García-Risco, L. Mouhid, L. Salas-Pérez, A. López-Padilla, S. Santoyo, L. Jaime, A. Ramírez de Molina, G. Reglero, T. Fornari, Biological activities of Asteraceae (Achillea millefolium and Calendula officinalis) and Lamiaceae (Melissa officinalis

and Origanum majorana) plant extracts, Plant Foods Hum. Nutr. 72 (2017) 96–102.

[195] B.P. Gupta, I. Sharma, N. Kohli, S. Sharma, A. Rathi, A.K. Sharma, Preliminary clinical assessment and non-toxicity evaluation of an ayurvedic formulation BGR-34 in NIDDM, J. Tradit. Complement. Med. 8 (2018) 506–514.

[196] Y. Yang, M.S. Islam, J. Wang, Y. Li, X. Chen, Traditional Chinese medicine in the treatment of patients infected with 2019-new coronavirus (SARS-CoV-2): a review and perspective, Int. J. Biol. Sci. 16 (2020) 1708.

[197] W. Chen, C.E.D. Lim, H.-J. Kang, J. Liu, Chinese herbal medicines for the treatment of type A H1N1 influenza: a systematic review of randomized controlled trials, PLoS One. 6 (2011) e28093.

[198] L. Xiaoyan, C.S. Lundborg, D. Banghan, C. Bojun, Z. Hong, L. Jiqiang, O. Aihua, O. Wenwei, W. Zehuai, L. Chuanjian, Clinical outcomes of influenza-like illness treated with Chinese herbal medicine: an observational study, J. Tradit. Chinese Med. 38 (2018) 107–116.

[199] L. Ang, H.W. Lee, J.Y. Choi, J. Zhang, M.S. Lee, Herbal medicine and pattern identification for treating COVID-19: a rapid review of guidelines, Integr. Med. Res. 9 (2020) 100407.

[200] J. Ding, B. Zhang, P.-J. Wang, G.-N. He, D.-M. Wei, J.-L. Ding, B.-C. Wang, Analysis on mechanisms and medication rules of herbal prescriptions for nonalcoholic fatty liver disease based on methods of data mining and biological information, Zhongguo Zhong Yao Za Zhi= Zhongguo Zhongyao Zazhi= China J. Chinese Mater. Medica. 44 (2019) 1689–1695.

[201] H. Luo, Q. Tang, Y. Shang, S. Liang, M. Yang, N. Robinson, J. Liu, Can Chinese medicine be used for prevention of corona virus disease 2019 (COVID-19)? A review of historical classics, research evidence and current prevention programs, Chin. J. Integr. Med. 26 (2020) 243–250.

[202] A.B. Gurung, M.A. Ali, J. Lee, M.A. Farah, K.M. Al-Anazi, An Updated Review of Computer-Aided Drug Design and Its Application to COVID-19, Biomed Res. Int. 2021 (2021) 8853056. https://doi.org/10.1155/2021/8853056.

[203] M. Laidi, A. Abdallah el Hadj, C. Si-Moussa, O. Benkortebi, M. Hentabli, S. Hanini, CMC of diverse Gemini surfactants modelling using a hybrid approach combining SVR-DA, Chem. Ind. Chem. Eng. Q. (2020) 48–48. https://doi.org/10.2298/ciceq200907048l.

[204] F. Omari, L. Khaouane, M. Laidi, A. Ibrir, M. Roubehie Fissa, M. Hentabli, S. Hanini, Dragonfly algorithm–support vector machine approach for prediction the optical properties of blood, Comput. Methods Biomech. Biomed. Engin. (2023) 1–10.

[205] N. Shaker, M. Abou-Zleikha, M. AlAmri, Y. Mehellou, A Generative deep learning approach for the discovery of SARS CoV2 protease inhibitors, (2020).

[206] I. Euldji, C. Si-Moussa, M. Hamadache, O. Benkortbi, QSPR Modelling of the Solubility of Drug and Drug-like Compounds in Supercritical Carbon Dioxide, Mol. Inform. 41 (2022) 2200026. https://doi.org/https://doi.org/10.1002/minf.202200026.

[207] M. Hamadache, O. Benkortbi, S. Hanini, A. Amrane, QSAR modeling in ecotoxicological risk assessment: application to the prediction of acute contact toxicity of pesticides on bees (Apis mellifera L.), Environ. Sci. Pollut. Res. 25 (2018) 896–907.

[208] M. Hamadache, O. Benkortbi, S. Hanini, A. Amrane, L. Khaouane, C. Si Moussa, A Quantitative Structure Activity Relationship for acute oral toxicity of pesticides on rats: Validation, domain of application and prediction, J. Hazard. Mater. 303 (2016) 28–40. https://doi.org/10.1016/j.jhazmat.2015.09.021.

[209] C.M. Song, S.J. Lim, J.C. Tong, Recent advances in computer-aided drug design, Brief. Bioinform. 10 (2009) 579–591.

[210] J.A. DiMasi, H.G. Grabowski, R.W. Hansen, Innovation in the pharmaceutical industry: new estimates of R&D costs, J. Health Econ. 47 (2016) 20–33.

[211] M. Hamadache, O. Benkortbi, S. Hanini, A. Amrane, Application of multilayer perceptron for prediction of the rat acute toxicity of insecticides, Energy Procedia. 139 (2017) 37–42. https://doi.org/10.1016/j.egypro.2017.11.169.

[212] S.J.Y. Macalino, V. Gosu, S. Hong, S. Choi, Role of computer-aided drug design in modern drug discovery, Arch. Pharm. Res. 38 (2015) 1686–1701.

[213] W. Duch, K. Swaminathan, J. Meller, Artificial intelligence approaches for rational drug design and discovery, Curr. Pharm. Des. 13 (2007) 1497–1508.

[214] T. Pillaiyar, S. Meenakshisundaram, M. Manickam, Recent discovery and development of inhibitors targeting coronaviruses, Drug Discov. Today. 25 (2020) 668–688.

[215] N.G. Bajad, S. Rayala, G. Gutti, A. Sharma, M. Singh, A. Kumar, S.K. Singh, Systematic review on role of structure based drug design (SBDD) in the identification of anti-viral leads against SARS-Cov-2, Curr. Res. Pharmacol. Drug Discov. 2 (2021) 100026. https://doi.org/https://doi.org/10.1016/j.crphar.2021.100026.

[216] S.-Y. Huang, X. Zou, Advances and challenges in protein-ligand docking, Int. J. Mol. Sci. 11 (2010) 3016–3034.

[217] I. Kapetanovic, Computer-aided drug discovery and development (CADDD): in silico-chemico-biological approach, Chem. Biol. Interact. 171 (2008) 165–176.

[218] S.F. Sousa, P.A. Fernandes, M.J. Ramos, Protein–ligand docking: current status and future challenges, Proteins Struct. Funct. Bioinforma. 65 (2006) 15–26.

[219] R. Dias, J. de Azevedo, F. Walter, Molecular docking algorithms, Curr. Drug Targets. 9 (2008) 1040–1047.

[220] K.K. Chaudhary, N. Mishra, A review on molecular docking: novel tool for drug discovery, Databases. 3 (2016) 1029.

[221] S. Agarwal, R. Mehrotra, An overview of molecular docking, JSM Chem. 4 (2016) 1024–1028.

[222] N.S. Pagadala, K. Syed, J. Tuszynski, Software for molecular docking: a review, Biophys. Rev. 9 (2017) 91–102.

[223] J. Fan, A. Fu, L. Zhang, Progress in molecular docking, Quant. Biol. 7 (2019) 83–89.

[224] E. Yuriev, P.A. Ramsland, Latest developments in molecular docking: 2010–2011 in review, J. Mol. Recognit. 26 (2013) 215–239.

[225] G.M. Morris, R. Huey, W. Lindstrom, M.F. Sanner, R.K. Belew, D.S. Goodsell, A.J. Olson, AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility, J. Comput. Chem. 30 (2009) 2785–2791.

[226] O. Trott, A.J. Olson, AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading, J. Comput. Chem. 31 (2010) 455–461. https://doi.org/https://doi.org/10.1002/jcc.21334.

[227] G. Jones, P. Willett, R.C. Glen, A.R. Leach, R. Taylor, Development and validation of a genetic algorithm for flexible docking, J. Mol. Biol. 267 (1997) 727–748.

[228] G. Wu, D.H. Robertson, C.L. Brooks III, M. Vieth, Detailed analysis of grid-based molecular docking: A case study of CDOCKER—A CHARMm-based MD docking algorithm, J. Comput. Chem. 24 (2003) 1549–1562.

[229] B. Kramer, M. Rarey, T. Lengauer, Evaluation of the FLEXX incremental construction algorithm for protein–ligand docking, Proteins Struct. Funct. Bioinforma. 37 (1999) 228–241.

[230] A.N. Jain, Surflex: fully automatic flexible molecular docking using a molecular similarity-based search engine, J. Med. Chem. 46 (2003) 499–511.

[231] R.A. Friesner, J.L. Banks, R.B. Murphy, T.A. Halgren, J.J. Klicic, D.T. Mainz, M.P. Repasky, E.H. Knoll, M. Shelley, J.K. Perry, Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy, J. Med. Chem. 47 (2004) 1739–1749.

[232] W.J. Allen, T.E. Balius, S. Mukherjee, S.R. Brozell, D.T. Moustakas, P.T. Lang, D.A. Case, I.D. Kuntz, R.C. Rizzo, DOCK 6: Impact of new features and current docking performance, J. Comput. Chem. 36 (2015) 1132–1156.

[233] A. Grosdidier, V. Zoete, O. Michielin, SwissDock, a protein-small molecule docking web service based on EADock DSS, Nucleic Acids Res. 39 (2011) W270–W277.

[234] C.C.G. Inc., MOE (The Molecular Operating Environment), (n.d.).

[235] G. Bitencourt-Ferreira, W.F. de Azevedo, Molegro Virtual Docker for Docking BT - Docking Screens for Drug Discovery, in: W.F. de Azevedo Jr. (Ed.), Springer New York, New York, NY, 2019: pp. 149–167. https://doi.org/10.1007/978-1-4939-9752-7_10.

[236] I. Canvas Schrödinger, Schrödinger Suite, (2015).

[237] X. Wang, K. Song, L. Li, L. Chen, Structure-based drug design strategies and challenges, Curr. Top. Med. Chem. 18 (2018) 998–1006.

[238] M.T. Muhammed, E. Aki-Yalcin, Homology modeling in drug discovery: Overview, current applications, and future perspectives, Chem. Biol. Drug Des. 93 (2019) 12–20.

[239] V.K. Vyas, R.D. Ukawala, M. Ghate, C. Chintha, Homology modeling a fast tool for drug discovery: current perspectives, Indian J. Pharm. Sci. 74 (2012) 1.

[240] J. Xu, F. Jiao, L. Yu, Protein structure prediction using threading, Protein Struct. Predict. (2008) 91–121.

[241] M. Yousef, T. Abdelkader, K. El-Bahnasy, Performance comparison of ab initio protein structure prediction methods, Ain Shams Eng. J. 10 (2019) 713–719.

[242] A. Raza, T.A. Chohan, M. Buabeid, E.-S.A. Arafa, T.A. Chohan, B. Fatima, K. Sultana, M.S. Ullah, G. Murtaza, Deep learning in drug discovery: a futuristic modality to materialize the large datasets for cheminformatics, J. Biomol. Struct. Dyn. (2022) 1–16. https://doi.org/10.1080/07391102.2022.2136244.

[243] V. Mody, J. Ho, S. Wills, A. Mawri, L. Lawson, M.C.C.J.C. Ebert, G.M. Fortin, S. Rayalam, S. Taval, Identification of 3-chymotrypsin like protease (3CLPro) inhibitors as potential anti-SARS-CoV-2 agents, Commun. Biol. 4 (2021). https://doi.org/10.1038/s42003-020-01577-x.

[244] M.C. Cobanoglu, C. Liu, F. Hu, Z.N. Oltvai, I. Bahar, Predicting drug-target interactions using probabilistic matrix factorization, J. Chem. Inf. Model. 53 (2013) 3399–3409. https://doi.org/10.1021/ci400219z.

[245] J.J. Irwin, B.K. Shoichet, ZINC− a free database of commercially available compounds for virtual screening, J. Chem. Inf. Model. 45 (2005) 177–182.

[246] S. Kim, P.A. Thiessen, E.E. Bolton, J. Chen, G. Fu, A. Gindulyte, L. Han, J. He, S. He, B.A. Shoemaker, PubChem substance and compound databases, Nucleic Acids Res. 44 (2016) D1202–D1213.

[247] H.E. Pence, A. Williams, ChemSpider: An Online Chemical Information Resource, J. Chem. Educ. 87 (2010) 1123–1124. https://doi.org/10.1021/ed100697w.

[248] D.S. Wishart, C. Knox, A.C. Guo, D. Cheng, S. Shrivastava, D. Tzur, B. Gautam, M. Hassanali, DrugBank: a knowledgebase for drugs, drug actions and drug targets, Nucleic Acids Res. 36 (2008) D901–D906.

[249] A. Gaulton, L.J. Bellis, A.P. Bento, J. Chambers, M. Davies, A. Hersey, Y. Light, S. McGlinchey, D. Michalovich, B. Al-Lazikani, ChEMBL: a large-scale bioactivity database for drug discovery, Nucleic Acids Res. 40 (2012) D1100–D1107.

[250] A. Daina, O. Michielin, V. Zoete, SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules, Sci. Rep. 7 (2017) 42717. https://doi.org/10.1038/srep42717.

[251] C.A. Lipinski, Lead-and drug-like compounds: the rule-of-five revolution, Drug Discov. Today Technol. 1 (2004) 337–341.

[252] J.A. McCammon, B.R. Gelin, M. Karplus, Dynamics of folded proteins, Nature. 267 (1977) 585–590.

[253] M. Karplus, J.A. McCammon, Molecular dynamics simulations of biomolecules, Nat. Struct. Biol. 9 (2002) 646–652.

[254] M. De Vivo, M. Masetti, G. Bottegoni, A. Cavalli, Role of molecular dynamics and related methods in drug discovery, J. Med. Chem. 59 (2016) 4035–4061.

[255] S.A. Hollingsworth, R.O. Dror, Molecular dynamics simulation for all, Neuron. 99 (2018) 1129–1143.

[256] D. Huang, A. Caflisch, The free energy landscape of small molecule unbinding, PLoS Comput. Biol. 7 (2011) e1002002.

[257] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A.E. Mark, H.J.C. Berendsen, GROMACS: fast, flexible, and free, J. Comput. Chem. 26 (2005) 1701–1718.

[258] D.A. Case, T.E. Cheatham III, T. Darden, H. Gohlke, R. Luo, K.M. Merz Jr, A. Onufriev, C. Simmerling, B. Wang, R.J. Woods, The Amber biomolecular simulation programs, J. Comput. Chem. 26 (2005) 1668–1688.

[259] B.R. Brooks, C.L. Brooks III, A.D. Mackerell Jr, L. Nilsson, R.J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, CHARMM: the biomolecular simulation program, J. Comput. Chem. 30 (2009) 1545–1614.

[260] J.C. Phillips, G. Zheng, S. Kumar, L. V Kalé, NAMD: Biomolecular simulation on thousands of processors, in: SC'02 Proc. 2002 ACM/IEEE Conf. Supercomput., IEEE, 2002: p. 36.

[261] J.A. Rackers, Z. Wang, C. Lu, M.L. Laury, L. Lagardère, M.J. Schnieders, J.-P. Piquemal, P. Ren, J.W. Ponder, Tinker 8: software tools for molecular design, J. Chem. Theory Comput. 14 (2018) 5273–5289.

[262] M. Karplus, G.A. Petsko, Molecular dynamics simulations in biology, Nature. 347 (1990) 631–639.

[263] M.P. Allen, Introduction to molecular dynamics simulation, Comput. Soft Matter from Synth. Polym. to Proteins. 23 (2004) 1–28.

[264] D.C. Rapaport, The art of molecular dynamics simulation, Cambridge university press, 2004.

[265] T. Hansson, C. Oostenbrink, W. van Gunsteren, Molecular dynamics simulations, Curr. Opin. Struct. Biol. 12 (2002) 190–196.

[266] A. Khaldan, S. Bouamrane, F. En-Nahli, R. El-mernissi, K. El khatabi, R. Hmamouchi, H. Maghat, M.A. Ajana, A. Sbai, M. Bouachrine, T. Lakhlifi, Prediction of potential inhibitors of SARS-CoV-2 using 3D-QSAR, molecular docking modeling and ADMET properties, Heliyon. 7 (2021). https://doi.org/10.1016/j.heliyon.2021.e06603.

[267] E.I. Edache, A. Uzairu, P.A. Mamza, G.A. Shallangwa, QSAR, homology modeling, and docking simulation on SARS-CoV-2 and pseudomonas aeruginosa inhibitors, ADMET, and molecular dynamic simulations to find a possible oral lead candidate., J. Genet. Eng. Biotechnol. 20 (2022) 88. https://doi.org/10.1186/s43141-022-00362-z.

[268] L.L.G. Ferreira, A.D. Andricopulo, ADMET modeling approaches in drug discovery, Drug Discov. Today. 24 (2019) 1157–1165. https://doi.org/https://doi.org/10.1016/j.drudis.2019.03.015.

[269] A. Aouidate, A. Ghaleb, S. Chtita, M. Aarjane, A. Ousaa, H. Maghat, A. Sbai, M.M. Choukrad, M. Bouachrine, T. Lakhlifi, Identification of a novel dual-target scaffold for 3CLpro and RdRp proteins of SARS-CoV-2 using 3D-similarity search, molecular

docking, molecular dynamics and ADMET evaluation., J. Biomol. Struct. Dyn. 39 (2021) 4522–4535. https://doi.org/10.1080/07391102.2020.1779130.

[270] D. Dey, R. Hossain, P. Biswas, P. Paul, M.A. Islam, T.I. Ema, B.K. Gain, M.M. Hasan, S. Bibi, M.T. Islam, M.A. Rahman, B. Kim, Amentoflavone derivatives significantly act towards the main protease (3CLPRO/MPRO) of SARS-CoV-2: in silico admet profiling, molecular docking, molecular dynamics simulation, network pharmacology, Mol. Divers. (2022). https://doi.org/10.1007/s11030-022-10459-9.

[271] G. Xiong, Z. Wu, J. Yi, L. Fu, Z. Yang, C. Hsieh, M. Yin, X. Zeng, C. Wu, A. Lu, X. Chen, T. Hou, D. Cao, ADMETlab 2.0: an integrated online platform for accurate and comprehensive predictions of ADMET properties., Nucleic Acids Res. 49 (2021) W5–W14. https://doi.org/10.1093/nar/gkab255.

[272] D.E. V Pires, T.L. Blundell, D.B. Ascher, pkCSM: Predicting Small-Molecule Pharmacokinetic and Toxicity Properties Using Graph-Based Signatures, J. Med. Chem. 58 (2015) 4066–4072. https://doi.org/10.1021/acs.jmedchem.5b00104.

[273] U. Norinder, C.A.S. Bergström, Prediction of ADMET properties, ChemMedChem Chem. Enabling Drug Discov. 1 (2006) 920–937.

[274] E.L. Cáceres, M. Tudor, A.C. Cheng, Deep learning approaches in predicting ADMET properties, Future Med. Chem. 12 (2020) 1995–1999.

[275] J.C. Dearden, In silico prediction of ADMET properties: how far have we come?, Expert Opin. Drug Metab. Toxicol. 3 (2007) 635–639.

[276] H. Van De Waterbeemd, E. Gifford, ADMET in silico modelling: towards prediction paradise?, Nat. Rev. Drug Discov. 2 (2003) 192–204.

[277] C.-Y. Jia, J.-Y. Li, G.-F. Hao, G.-F. Yang, A drug-likeness toolbox facilitates ADMET study in drug discovery, Drug Discov. Today. 25 (2020) 248–258.

[278] D.E. Clark, P.D. Grootenhuis, Progress in computational methods for the prediction of ADMET properties., Curr. Opin. Drug Discov. Devel. 5 (2002) 382–390.

[279] F. Cheng, W. Li, G. Liu, Y. Tang, In silico ADMET prediction: recent advances, current challenges and future trends, Curr. Top. Med. Chem. 13 (2013) 1273–1289.

[280] J. Gola, O. Obrezanova, E. Champness, M. Segall, ADMET property prediction: the state of the art and current challenges, QSAR Comb. Sci. 25 (2006) 1172–1180.

[281] A.M. Davis, R.J. Riley, Predictive ADMET studies, the challenges and the opportunities, Curr. Opin. Chem. Biol. 8 (2004) 378–386.

[282] S. Das, S. Sarmah, S. Lyndem, A. Singha Roy, An investigation into the identification of potential inhibitors of SARS-CoV-2 main protease using molecular docking study, J. Biomol. Struct. Dyn. 39 (2021) 3347–3357. https://doi.org/10.1080/07391102.2020.1763201.

[283] R.R. Deshpande, A.P. Tiwari, N. Nyayanit, M. Modak, In silico molecular docking analysis for repurposing therapeutics against multiple proteins from SARS-CoV-2, Eur. J. Pharmacol. 886 (2020) 173430. https://doi.org/https://doi.org/10.1016/j.ejphar.2020.173430.

[284] R. kerkour, N. Chafai, O. Moumeni, S. Chafaa, Novel α-aminophosphonate derivates synthesis, theoretical calculation, Molecular docking, and in silico prediction of potential inhibition of SARS-CoV-2, J. Mol. Struct. 1272 (2023) 134196. https://doi.org/https://doi.org/10.1016/j.molstruc.2022.134196.

[285] S. Mhatre, S. Naik, V. Patravale, A molecular docking study of EGCG and theaflavin digallate with the druggable targets of SARS-CoV-2, Comput. Biol. Med. 129 (2021) 104137. https://doi.org/https://doi.org/10.1016/j.compbiomed.2020.104137.

[286] R.P. Vivek-Ananth, S. Krishnaswamy, A. Samal, Potential phytochemical inhibitors of SARS-CoV-2 helicase Nsp13: a molecular docking and dynamic simulation study, Mol. Divers. 26 (2022) 429–442. https://doi.org/10.1007/s11030-021-10251-1.

[287] R. Ghosh, A. Chakraborty, A. Biswas, S. Chowdhuri, Evaluation of green tea polyphenols as novel corona virus (SARS CoV-2) main protease (Mpro) inhibitors – an in silico docking and molecular dynamics simulation study, J. Biomol. Struct. Dyn. 39 (2021) 4362–4374. https://doi.org/10.1080/07391102.2020.1779818.

[288] D.C. Hall, H.-F. Ji, A search for medications to treat COVID-19 via in silico molecular docking models of the SARS-CoV-2 spike glycoprotein and 3CL protease, Travel Med. Infect. Dis. 35 (2020) 101646. https://doi.org/https://doi.org/10.1016/j.tmaid.2020.101646.

[289] D. Li, J. Luan, L. Zhang, Molecular docking of potential SARS-CoV-2 papain-like protease inhibitors, Biochem. Biophys. Res. Commun. 538 (2021) 72–79. https://doi.org/https://doi.org/10.1016/j.bbrc.2020.11.083.

[290] F. Rahman, S. Tabrez, R. Ali, A.S. Alqahtani, M.Z. Ahmed, A. Rub, Molecular docking analysis of rutin reveals possible inhibition of SARS-CoV-2 vital proteins, J. Tradit. Complement. Med. 11 (2021) 173–179. https://doi.org/https://doi.org/10.1016/j.jtcme.2021.01.006.

[291] K. El Khatabi, I. Aanouz, M. Alaqarbeh, M.A. Ajana, T. Lakhlifi, M. Bouachrine, Molecular docking, molecular dynamics simulation, and ADMET analysis of levamisole derivatives against the SARS-CoV-2 main protease (M(Pro))., Bioimpacts. 12 (2022) 107–113. https://doi.org/10.34172/bi.2021.22143.

[292] M.I. Choudhary, M. Shaikh, A.- tul-Wahab, A.- ur-Rahman, In silico identification of potential inhibitors of key SARS-CoV-2 3CL hydrolase (Mpro) via molecular docking, MMGBSA predictive binding energy calculations, and molecular dynamics simulation, PLoS One. 15 (2020) e0235030. https://doi.org/10.1371/journal.pone.0235030.

[293] H.M. Wahedi, S. Ahmad, S.W. Abbasi, Stilbene-based natural compounds as promising drug candidates against COVID-19, J. Biomol. Struct. Dyn. 39 (2021) 3225–3234. https://doi.org/10.1080/07391102.2020.1762743.

[294] A.A. Al-Karmalawy, M.A. Dahab, A.M. Metwaly, S.S. Elhady, E.B. Elkaeed, I.H. Eissa, K.M. Darwish, Molecular Docking and Dynamics Simulation Revealed the Potential Inhibitory Activity of ACEIs Against SARS-CoV-2 Targeting the hACE2 Receptor , Front. Chem. . 9 (2021). https://www.frontiersin.org/articles/10.3389/fchem.2021.661230.

[295] P.P. Kushwaha, A.K. Singh, T. Bansal, A. Yadav, K.S. Prajapati, M. Shuaib, S. Kumar,

Identification of Natural Inhibitors Against SARS-CoV-2 Drugable Targets Using Molecular Docking, Molecular Dynamics Simulation, and MM-PBSA Approach , Front. Cell. Infect. Microbiol. . 11 (2021). https://www.frontiersin.org/articles/10.3389/fcimb.2021.730288.

[296] S.K. Enmozhi, K. Raja, I. Sebastine, J. Joseph, Andrographolide as a potential inhibitor of SARS-CoV-2 main protease: an in silico approach, J. Biomol. Struct. Dyn. 39 (2021) 3092–3098. https://doi.org/10.1080/07391102.2020.1760136.

[297] S.A. Khan, K. Zia, S. Ashraf, R. Uddin, Z. Ul-Haq, Identification of chymotrypsin-like protease inhibitors of SARS-CoV-2 via integrated computational approach, J. Biomol. Struct. Dyn. 39 (2021) 2607–2616.

[298] N.D. Gajjar, T.M. Dhameliya, G.B. Shah, In search of RdRp and Mpro inhibitors against SARS CoV-2: Molecular docking, molecular dynamic simulations and ADMET analysis, J. Mol. Struct. 1239 (2021) 130488. https://doi.org/https://doi.org/10.1016/j.molstruc.2021.130488.

[299] M. Rudrapal, N. Gogoi, D. Chetia, J. Khan, S. Banwas, B. Alshehri, M.A. Alaidarous, U.D. Laddha, S.J. Khairnar, S.G. Walode, Repurposing of phytomedicine-derived bioactive compounds with promising anti-SARS-CoV-2 potential: Molecular docking, MD simulation and drug-likeness/ADMET studies, Saudi J. Biol. Sci. 29 (2022) 2432–2446. https://doi.org/https://doi.org/10.1016/j.sjbs.2021.12.018.

[300] S. M. A. Kawsar, M.A. Hosen, S. Ahmad, Y. El Bakri, H. Laaroussi, T. Ben Hadda, F.A. Almalki, Y. Ozeki, S. Goumri-Said, Potential SARS-CoV-2 RdRp inhibitors of cytidine derivatives: Molecular docking, molecular dynamic simulations, ADMET, and POM analyses for the identification of pharmacophore sites, PLoS One. 17 (2022) e0273256. https://doi.org/10.1371/journal.pone.0273256.

[301] A. Belhassan, S. Chtita, H. Zaki, M. Alaqarbeh, N. Alsakhen, F. Almohtaseb, T. Lakhlifi, M. Bouachrine, In silico detection of potential inhibitors from vitamins and their derivatives compounds against SARS-CoV-2 main protease by using molecular docking, molecular dynamic simulation and ADMET profiling, J. Mol. Struct. 1258 (2022) 132652. https://doi.org/https://doi.org/10.1016/j.molstruc.2022.132652.

[302] H. Hadni, A. Fitri, A.T. Benjelloun, M. Benzakour, M. Mcharfi, Evaluation of flavonoids as potential inhibitors of the SARS-CoV-2 main protease and spike RBD: Molecular docking, ADMET evaluation and molecular dynamics simulations, J. Indian Chem. Soc. 99 (2022) 100697. https://doi.org/https://doi.org/10.1016/j.jics.2022.100697.

[303] D. Schaller, D. Šribar, T. Noonan, L. Deng, T.N. Nguyen, S. Pach, D. Machalz, M. Bermudez, G. Wolber, Next generation 3D pharmacophore modeling, Wiley Interdiscip. Rev. Comput. Mol. Sci. 10 (2020) e1468.

[304] J.H. Van Drie, Generation of three-dimensional pharmacophore models, Wiley Interdiscip. Rev. Comput. Mol. Sci. 3 (2013) 449–464.

[305] C.C. Melo-Filho, R.C. Braga, C.H. Andrade, 3D-QSAR approaches in drug design: perspectives to generate reliable CoMFA models., Curr. Comput. Aided. Drug Des. 10 (2014) 148–159.

[306] A. Vuorinen, D. Schuster, Methods for generating and applying pharmacophore models as virtual screening filters and for bioactivity profiling, Methods. 71 (2015) 113–134.

[307] T. Klabunde, C. Giegerich, A. Evers, Sequence-derived three-dimensional pharmacophore models for G-protein-coupled receptors and their application in virtual screening, J. Med. Chem. 52 (2009) 2923–2932.

[308] M.J. de Groot, S. Ekins, Pharmacophore modeling of cytochromes P450, Adv. Drug Deliv. Rev. 54 (2002) 367–383.

[309] M. Wieder, A. Garon, U. Perricone, S. Boresch, T. Seidel, A.M. Almerico, T. Langer, Common hits approach: combining pharmacophore modeling and molecular dynamics simulations, J. Chem. Inf. Model. 57 (2017) 365–385.

[310] T. Kaserer, K.R. Beck, M. Akram, A. Odermatt, D. Schuster, Pharmacophore models and pharmacophore-based virtual screening: concepts and applications exemplified on hydroxysteroid dehydrogenases, Molecules. 20 (2015) 22799–22832.

[311] D. Schneidman-Duhovny, O. Dror, Y. Inbar, R. Nussinov, H.J. Wolfson, PharmaGist: a webserver for ligand-based pharmacophore detection, Nucleic Acids Res. 36 (2008) W223–W228.

[312] X. Liu, S. Ouyang, B. Yu, Y. Liu, K. Huang, J. Gong, S. Zheng, Z. Li, H. Li, H. Jiang, PharmMapper server: a web server for potential drug target identification using pharmacophore mapping approach, Nucleic Acids Res. 38 (2010) W609–W614.

[313] D.R. Koes, C.J. Camacho, Pharmer: efficient and exact pharmacophore search, J. Chem. Inf. Model. 51 (2011) 1307–1314.

[314] S.-Y. Yang, Pharmacophore modeling and applications in drug discovery: challenges and recent advances, Drug Discov. Today. 15 (2010) 444–450.

[315] S.A. Khedkar, A.K. Malde, E.C. Coutinho, S. Srivastava, Pharmacophore modeling in drug discovery and development: an overview, Med. Chem. (Los. Angeles). 3 (2007) 187–197.

[316] X. Qing, X. Yin Lee, J. De Raeymaeker, J. RH Tame, K. YJ Zhang, M. De Maeyer, A. RD Voet, Pharmacophore modeling: advances, limitations, and current utility in drug discovery, J. Receptor. Ligand Channel Res. (2014) 81–92.

[317] Y. Kurogi, O.F. Guner, Pharmacophore modeling and three-dimensional database searching for drug design using catalyst, Curr. Med. Chem. 8 (2001) 1035–1055.

[318] D. Barnum, J. Greene, A. Smellie, P. Sprague, Identification of common functional configurations among molecules, J. Chem. Inf. Comput. Sci. 36 (1996) 563–571.

[319] G. Wolber, T. Langer, LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters, J. Chem. Inf. Model. 45 (2005) 160–169.

[320] A. Cherkasov, E.N. Muratov, D. Fourches, A. Varnek, I.I. Baskin, M. Cronin, J. Dearden, P. Gramatica, Y.C. Martin, R. Todeschini, QSAR modeling: where have you been? Where are you going to?, J. Med. Chem. 57 (2014) 4977–5010.

[321] K. Roy, S. Kar, R.N. Das, Chapter 7 - Validation of QSAR Models, in: K. Roy, S. Kar, R.N.B.T.-U. the B. of Q. for A. in P.S. and R.A. Das (Eds.), Academic Press, Boston, 2015: pp. 231–289. https://doi.org/https://doi.org/10.1016/B978-0-12-801505-6.00007-7.

[322] K. Roy, S. Kar, R. Das, A primer on QSAR/QSPR modeling: fundamental concepts, 2015. https://doi.org/10.1007/978-3-319-17281-1.

[323] K. Roy, I. Mitra, Electrotopological State Atom (E-State) Index in Drug Design, QSAR, Property Prediction and Toxicity Assessment, Curr. Comput. Aided-Drug Des. 8 (2012) 135–158. https://doi.org/10.2174/157340912800492366.

[324] A. Mauri, alvaDesc: A Tool to Calculate and Analyze Molecular Descriptors and Fingerprints BT - Ecotoxicological QSARs, in: K. Roy (Ed.), Springer US, New York, NY, 2020: pp. 801–820. https://doi.org/10.1007/978-1-0716-0150-1_32.

[325] J. Dong, D.S. Cao, H.Y. Miao, S. Liu, B.C. Deng, Y.H. Yun, N.N. Wang, A.P. Lu, W. Bin Zeng, A.F. Chen, ChemDes: An integrated web-based platform for molecular descriptor and fingerprint computation, J. Cheminform. 7 (2015) 1–10. https://doi.org/10.1186/s13321-015-0109-z.

[326] T. Fe, Electronic Supplementary Information, Scanning. (2011) 1–10.

[327] C.W. Yap, PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints, J. Comput. Chem. 32 (2011) 1466–1474.

[328] H. Li, J. Sutter, R. Hoffmann, HypoGen: an automated system for generating 3D predictive pharmacophore models, Pharmacophore Perception, Dev. Use Drug Des. 2 (2000) 171.

[329] S.L. Dixon, A.M. Smondyrev, S.N. Rao, PHASE: a novel approach to pharmacophore modeling and 3D database searching, Chem. Biol. Drug Des. 67 (2006) 370–372.

[330] G. Klebe, U. Abraham, T. Mietzner, Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity, J. Med. Chem. 37 (1994) 4130–4146.

[331] M.T.D. Cronin, T.W. Schultz, Pitfalls in QSAR, J. Mol. Struct. THEOCHEM. 622 (2003) 39–51.

[332] P. Gramatica, Principles of QSAR models validation: internal and external, QSAR Comb. Sci. 26 (2007) 694–701.

[333] E.N. Muratov, J. Bajorath, R.P. Sheridan, I. V Tetko, D. Filimonov, V. Poroikov, T.I. Oprea, I.I. Baskin, A. Varnek, A. Roitberg, QSAR without borders, Chem. Soc. Rev. 49 (2020) 3525–3564.

[334] K. Roy, P. Ambure, The “double cross-validation” software tool for MLR QSAR model development, Chemom. Intell. Lab. Syst. 159 (2016) 108–126. https://doi.org/https://doi.org/10.1016/j.chemolab.2016.10.009.

[335] P.K. Ojha, I. Mitra, R.N. Das, K. Roy, Further exploring rm2 metrics for validation of QSPR models, Chemom. Intell. Lab. Syst. 107 (2011) 194–205. https://doi.org/https://doi.org/10.1016/j.chemolab.2011.03.011.

[336] K. Roy, R.N. Das, P. Ambure, R.B. Aher, Be aware of error measures. Further studies on validation of predictive QSAR models, Chemom. Intell. Lab. Syst. 152 (2016) 18–33. https://doi.org/10.1016/J.CHEMOLAB.2016.01.008.

[337] K. Roy, P. Ambure, R.B. Aher, How important is to detect systematic error in predictions and understand statistical applicability domain of QSAR models?, Chemom. Intell. Lab. Syst. 162 (2017) 44–54.

[338] A.Z. Dudek, T. Arodz, J. Gálvez, Computational methods in developing quantitative structure-activity relationships (QSAR): a review, Comb. Chem. High Throughput Screen. 9 (2006) 213–228.

[339] J. Verma, V.M. Khedkar, E.C. Coutinho, 3D-QSAR in drug design-a review, Curr. Top. Med. Chem. 10 (2010) 95–115.

[340] H.M. Patel, M.N. Noolvi, P. Sharma, V. Jaiswal, S. Bansal, S. Lohan, S.S. Kumar, V. Abbot, S. Dhiman, V. Bhardwaj, Quantitative structure–activity relationship (QSAR) studies as strategic approach in drug discovery, Med. Chem. Res. 23 (2014) 4991–5007.

[341] L. Wang, J. Ding, L. Pan, D. Cao, H. Jiang, X. Ding, Artificial intelligence facilitates drug design in the big data era, Chemom. Intell. Lab. Syst. 194 (2019) 103850.

[342] L. Patel, T. Shukla, X. Huang, D.W. Ussery, S. Wang, Machine learning methods in drug discovery, Molecules. 25 (2020) 5277.

[343] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32.

[344] G. Webb, Naïve Bayes, in: 2016: pp. 1–2. https://doi.org/10.1007/978-1-4899-7502-7_581-1.

[345] F. Zhong, J. Xing, X. Li, X. Liu, Z. Fu, Z. Xiong, D. Lu, X. Wu, J. Zhao, X. Tan, Artificial intelligence in drug design, Sci. China Life Sci. 61 (2018) 1191–1204.

[346] P.P. Roy, J.T. Leonard, K. Roy, Exploring the impact of size of training sets for the development of predictive QSAR models, Chemom. Intell. Lab. Syst. 90 (2008) 31–42.

[347] T.R. Stouch, J.R. Kenyon, S.R. Johnson, X.-Q. Chen, A. Doweyko, Y. Li, In silico ADME/Tox: why models fail, J. Comput. Aided. Mol. Des. 17 (2003) 83–92.

[348] H. Zhu, M.T. Pisabarro, MSPocket: an orientation-independent algorithm for the detection of ligand binding pockets, Bioinformatics. 27 (2011) 351–358.

[349] P. Probst, A.L. Boulesteix, B. Bischl, Tunability: Importance of hyperparameters of machine learning algorithms, J. Mach. Learn. Res. 20 (2019) 1–32. https://doi.org/https://doi.org/10.48550/arXiv.1802.09596.

[350] Q. Shen, J.H. Jiang, C.X. Jiao, G.L. Shen, R.Q. Yu, Modified particle swarm optimization algorithm for variable selection in MLR and PLS modeling: QSAR studies of antagonism of angiotensin II antagonists, Eur. J. Pharm. Sci. 22 (2004) 145–152. https://doi.org/10.1016/j.ejps.2004.03.002.

[351] A. Kumar, Chapter 5 - Application of nature-inspired computing paradigms in optimal design of structural engineering problems—a review, in: M.A. Mellal, M.G.B.T.-N.-

I.C.P. in S. Pecht (Eds.), Intell. Data-Centric Syst., Academic Press, 2021: pp. 63–74. https://doi.org/https://doi.org/10.1016/B978-0-12-823749-6.00010-6.

[352] S. Mirjalili, Dragonfly algorithm: a new meta-heuristic optimization technique for solving single-objective, discrete, and multi-objective problems, Neural Comput. Appl. 27 (2016) 1053–1073.

[353] J. Kennedy, R. Eberhart, Particle swarm optimization, in: Proc. ICNN'95 - Int. Conf. Neural Networks, 1995: pp. 1942–1948 vol.4. https://doi.org/10.1109/ICNN.1995.488968.

[354] M. Dorigo, M. Birattari, T. Stutzle, Ant colony optimization, IEEE Comput. Intell. Mag. 1 (2006) 28–39.

[355] R.A. Rutenbar, Simulated annealing algorithms: An overview, IEEE Circuits Devices Mag. 5 (1989) 19–26.

[356] S. Mirjalili, S.M. Mirjalili, A. Lewis, Grey wolf optimizer, Adv. Eng. Softw. 69 (2014) 46–61.

[357] S. Mirjalili, A. Lewis, The whale optimization algorithm, Adv. Eng. Softw. 95 (2016) 51–67.

[358] A. Ibrir, Y. Kerchich, N. Hadidi, H. Merabet, M. Hentabli, Prediction of the concentrations of PM1, PM2.5, PM4, and PM10 by using the hybrid dragonfly-SVM algorithm, Air Qual. Atmos. Heal. 14 (2021) 313–323. https://doi.org/10.1007/s11869-020-00936-1.

[359] Y. Meraihi, A. Ramdane-Cherif, D. Acheli, M. Mahseur, Dragonfly algorithm: a comprehensive review and applications, Neural Comput. Appl. 32 (2020) 16625–16646. https://doi.org/10.1007/s00521-020-04866-y.

[360] F.D. Wihartiko, H. Wijayanti, F. Virgantari, Performance comparison of genetic algorithms and particle swarm optimization for model integer programming bus timetabling problem, IOP Conf. Ser. Mater. Sci. Eng. 332 (2018) 012020. https://doi.org/10.1088/1757-899X/332/1/012020.

[361] H. Alkabbani, A. Ahmadian, Q. Zhu, A. Elkamel, Machine Learning and Metaheuristic Methods for Renewable Power Forecasting: A Recent Review, Front. Chem. Eng. 3 (2021) 1–21. https://doi.org/10.3389/fceng.2021.665415.

[362] V. Rotemberg, A. Halpern, S. Dusza, N.C.F. Codella, The role of public challenges and data sets towards algorithm development, trust, and use in clinical practice., in: Semin. Cutan. Med. Surg., 2019: pp. E38–E42.

[363] M. Davies, M. Nowotka, G. Papadatos, N. Dedman, A. Gaulton, F. Atkinson, L. Bellis, J.P. Overington, ChEMBL web services: streamlining access to drug discovery data and utilities, Nucleic Acids Res. 43 (2015) W612–W620.

[364] X. Yu, Prediction of inhibitory constants of compounds against SARS-CoV 3CLpro enzyme with 2D-QSAR model, J. Saudi Chem. Soc. 25 (2021) 101–262. https://doi.org/https://doi.org/10.1016/j.jscs.2021.101262.

[365] G.I.B. Janairo, D.E.C. Yu, J.I.B. Janairo, A machine learning regression model for the

screening and design of potential SARS-CoV-2 protease inhibitors, Netw. Model. Anal. Heal. Informatics Bioinforma. 10 (2021) 1–8. https://doi.org/10.1007/s13721-021-00326-2.

[366] A.A. Ishola, O. Adedirin, T. Joshi, S. Chandra, QSAR modeling and pharmacoinformatics of SARS coronavirus 3C-like protease inhibitors, Comput. Biol. Med. 134 (2021) 104–483. https://doi.org/https://doi.org/10.1016/j.compbiomed.2021.104483.

[367] I. Hammoudan, S. Matchi, M. Bakhouch, S. Belaidi, S. Chtita, QSAR Modelling of Peptidomimetic Derivatives towards HKU4-CoV 3CL pro Inhibitors against MERS-CoV, (2021) 391–401.

[368] Z. Mozafari, M.A. Chamjangali, M. Arashi, N. Goudarzi, Suggestion of active 3-chymotrypsin like protease (3CL Pro) inhibitors as potential anti-SARS-CoV-2 agents using predictive QSAR model based on the combination of ALASSO with an ANN model, SAR QSAR Environ. Res. 32 (2021) 863–888. https://doi.org/10.1080/1062936X.2021.1975167.

[369] A.A. Toropov, A.P. Toropova, A.M. Veselinović, D. Leszczynska, J. Leszczynski, SARS-CoV Mpro inhibitory activity of aromatic disulfide compounds: QSAR model., J. Biomol. Struct. Dyn. 40 (2022) 780–786. https://doi.org/10.1080/07391102.2020.1818627.

[370] O.O. Elekofehinti, O. Iwaloye, O.R. Molehin, C.D. Famusiwa, Identification of lead compounds from large natural product library targeting 3C-like protease of SARS-CoV-2 using E-pharmacophore modelling, QSAR and molecular dynamics simulation, Silico Pharmacol. 9 (2021) 1–19. https://doi.org/10.1007/s40203-021-00109-7.

[371] N. Adhikari, S. Banerjee, S.K. Baidya, B. Ghosh, T. Jha, Ligand-based quantitative structural assessments of SARS-CoV-2 3CLpro inhibitors: An analysis in light of structure-based multi-molecular modeling evidences, J. Mol. Struct. 1251 (2022) 132041. https://doi.org/10.1016/j.molstruc.2021.132041.

[372] O. Chedadi, A. Aissouq, Y. El Ouardi, M. Bouachrine, O. Abdelkrim, In silico Prediction of Novel SARS-CoV 3CL pro Inhibitors: a Combination of 3D-QSAR, Molecular Docking, ADMET Prediction, and Molecular Dynamics Simulation, (2021). https://doi.org/10.33263/BRIAC124.51005115.

[373] K. Gholivand, A. Barzegari, F. Mohammadpanah, R. Yaghoubi, R. Roohzadeh, A.A. Ebrahimi Valmoozi, Synthesis, characterized, QSAR studies and molecular docking of some phosphonates as COVID-19 inhibitors, Polyhedron. 221 (2022) 115824. https://doi.org/https://doi.org/10.1016/j.poly.2022.115824.

[374] M.A. Khanfar, Structure-based Pharmacophore Screening Coupled with QSAR Analysis Identified Potent Natural-product-derived IRAK-4 Inhibitors, Mol. Inform. 40 (2021) 2100025. https://doi.org/https://doi.org/10.1002/minf.202100025.

[375] S. Saha, D.K. Pal, 2d Qsar Approach to Develop Newer Generation Small Molecules Active against Sars-Covid, Int. J. Pharm. Sci. Res. (2023) 1372–1391.

[376] N. Soleymani, S. Ahmadi, F. Shiri, A. Almasirad, QSAR and molecular docking studies of isatin and indole derivatives as SARS 3CLpro inhibitors, BMC Chem. 17

(2023) 32. https://doi.org/10.1186/s13065-023-00947-w.

[377] S.A. Amin, K. Ghosh, S. Gayen, T. Jha, Chemical-informatics approach to COVID-19 drug discovery: Monte Carlo based QSAR, virtual screening and molecular docking study of some in-house molecules as papain-like protease (PLpro) inhibitors, J. Biomol. Struct. Dyn. 39 (2021) 4764–4773.

[378] K. Ghosh, S.A. Amin, S. Gayen, T. Jha, Chemical-informatics approach to COVID-19 drug discovery: Exploration of important fragments and data mining based prediction of some hits from natural origins as main protease (Mpro) inhibitors, J. Mol. Struct. 1224 (2021) 129026. https://doi.org/https://doi.org/10.1016/j.molstruc.2020.129026.

[379] W.Y. Law, M.R. Asaruddin, S.A. Bhawani, S. Mohamad, Pharmacophore modelling of vanillin derivatives, favipiravir, chloroquine, hydroxychloroquine, monolaurin and tetrodotoxin as M pro inhibitors of severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2), BMC Res. Notes. 13 (2020) 1–8.

[380] M.O. Rafi, G. Bhattacharje, K. Al-Khafaji, T. Taskin-Tok, M.A. Alfasane, A.K. Das, M.A.K. Parvez, M.S. Rahman, Combination of QSAR, molecular docking, molecular dynamic simulation and MM-PBSA: analogues of lopinavir and favipiravir as potential drug candidates against COVID-19, J. Biomol. Struct. Dyn. 40 (2022) 3711–3730. https://doi.org/10.1080/07391102.2020.1850355.

[381] B. Thorat, V.P. Purohit, R. Yamgar, D. Bhagat, S.D. Wavhal, S.N. Mali, Structural Insight into 2-Aryl-4-Quinoline Carboxylic Acid-Based Dihydroorotate Dehydrogenase (DHODH) and its Potential Anti-SARS-CoV-2 Activity Through Pharmacophore Modeling, Multidimensional QSAR, ADME, and Docking Studies, Phys. Chem. Res. 11 (2023) 783–800. https://doi.org/10.22036/pcr.2022.365408.2218.

[382] A.S. Reddy, S.P. Pati, P.P. Kumar, H.N. Pradeep, G.N. Sastry, Virtual screening in drug discovery -- a computational perspective., Curr. Protein Pept. Sci. 8 (2007) 329–351. https://doi.org/10.2174/138920307781369427.

[383] T. Hou, X. Xu, Recent development and application of virtual screening in drug discovery: an overview, Curr. Pharm. Des. 10 (2004) 1011–1033.

[384] J. Alvarez, B. Shoichet, Virtual screening in drug discovery, CRC press, 2005.

[385] M. Kontoyianni, Docking and virtual screening in drug discovery, Proteomics Drug Discov. Methods Protoc. (2017) 255–266.

[386] T. Cheng, Q. Li, Z. Zhou, Y. Wang, S.H. Bryant, Structure-based virtual screening for drug discovery: a problem-centric review, AAPS J. 14 (2012) 133–141.

[387] A.R. Patel, H.B. Patel, S.K. Mody, R.D. Singh, V.N. Sarvaiya, S.H. Vaghela, S. Tukra, Virtual screening in drug discovery, J. Vet. Pharmacol. Toxicol. 20 (2021) 1–9.

[388] S.H. Han, C.M. Goins, T. Arya, W.J. Shin, J. Maw, A. Hooper, D.P. Sonawane, M.R. Porter, B.E. Bannister, R.D. Crouch, A.A. Lindsey, G. Lakatos, S.R. Martinez, J. Alvarado, W.S. Akers, N.S. Wang, J.U. Jung, J.D. MacDonald, S.R. Stauffer, Structure-Based Optimization of ML300-Derived, Noncovalent Inhibitors Targeting the Severe Acute Respiratory Syndrome Coronavirus 3CL Protease (SARS-CoV-2 3CLpro), J. Med. Chem. 4 (2021) 2880–2904. https://doi.org/10.1021/acs.jmedchem.1c00598.

[389] L. Fu, F. Ye, Y. Feng, F. Yu, Q. Wang, Y. Wu, C. Zhao, H. Sun, B. Huang, P. Niu, H. Song, Y. Shi, X. Li, W. Tan, J. Qi, G.F. Gao, Both Boceprevir and GC376 efficaciously inhibit SARS-CoV-2 by targeting its main protease, Nat. Commun. 11 (2020) 44–17. https://doi.org/10.1038/s41467-020-18233-x.

[390] A. Aljoundi, I. Bjij, A. Rashedy, M. Soliman, Covalent Versus Non-covalent Enzyme Inhibition: Which Route Should We Take? A Justification of the Good and Bad from Molecular Modelling Perspective, Protein J. 39 (2020) 97–105. https://doi.org/10.1007/s10930-020-09884-2.

[391] S.E.S. John, S. Tomar, S.R. Stauffer, A.D. Mesecar, Targeting zoonotic viruses: Structure-based inhibition of the 3C-like protease from bat coronavirus HKU4—The likely reservoir host to the human coronavirus that causes Middle East Respiratory Syndrome (MERS), Bioorg. Med. Chem. 23 (2015) 6036–6048. https://doi.org/10.1016/j.bmc.2015.06.039.

[392] T. Sander, J. Freyss, M. von Korff, C. Rufener, DataWarrior: An Open-Source Program For Chemistry Aware Data Visualization And Analysis, J. Chem. Inf. Model. 55 (2015) 460–473. https://doi.org/10.1021/ci500588j.

[393] G.K. Uyanık, N. Güler, A Study on Multiple Linear Regression Analysis, Procedia - Soc. Behav. Sci. 106 (2013) 234–240. https://doi.org/10.1016/j.sbspro.2013.12.027.

[394] M. Awad, R. Khanna, Support Vector Regression, in: Effic. Learn. Mach. Theor. Concepts, Appl. Eng. Syst. Des., Apress, Berkeley, CA, CA, 2015: pp. 67–80. https://doi.org/10.1007/978-1-4302-5990-9_4.

[395] M. Rupp, Machine learning for quantum mechanics in a nutshell, Int. J. Quantum Chem. 115 (2015) 1058–1073. https://doi.org/https://doi.org/10.1002/qua.24954.

[396] B. Liu, C. Jin, J. Wan, P. Li, H. Yan, Modelling and optimizing an electrochemical oxidation process using artificial neural network, genetic algorithm and particle swarm optimization, J. Serbian Chem. Soc. 83 (2018) 379–390. https://doi.org/10.2298/JSC170721101L.

[397] B.S.G. de Almeida, Particle Swarm Optimization: A Powerful Technique for Solving Engineering Problems, in: V.C.L.E.-J.D.S.E.-E.V.E.-E. Osaba (Ed.), IntechOpen, Rijeka, 2019: p. Ch. 3. https://doi.org/10.5772/intechopen.89633.

[398] P. Mitikiri, G. Jana, S. Sural, P.K. Chattaraj, A machine learning technique toward generating minimum energy structures of small boron clusters, Int. J. Quantum Chem. 118 (2018) 1–15. https://doi.org/https://doi.org/10.1002/qua.25672.

[399] O. Falyouna, O. Eljamal, I. Maamoun, A. Tahara, Y. Sugihara, Magnetic zeolite synthesis for efficient removal of cesium in a lab-scale continuous treatment system, J. Colloid Interface Sci. 571 (2020) 66–79. https://doi.org/https://doi.org/10.1016/j.jcis.2020.03.028.

[400] V. Suresh, S. Sreejith, Generation dispatch of combined solar thermal systems using dragonfly algorithm, Computing. 99 (2017) 59–80.

[401] J.D. Wichard, M.J. Ogorzałek, C. Merkwirth, CNN in drug design — Recent developments, in: 2015 IEEE Int. Symp. Circuits Syst., 2015: pp. 405–408. https://doi.org/10.1109/ISCAS.2015.7168656.

[402] E. Sariev, G. Germano, Bayesian regularized artificial neural networks for the estimation of the probability of default, Quant. Financ. 20 (2020) 311–328. https://doi.org/10.1080/14697688.2019.1633014.

[403] F. Bin Ashraf, S. Akter, S.H. Mumu, M.U. Islam, J. Uddin, Bio-activity prediction of drug candidate compounds targeting SARS-Cov-2 using machine learning approaches, PLoS One. 18 (2023) e0288053. https://doi.org/10.1371/journal.pone.0288053.

[404] E. Johansson, L. Eriksson, M. Sandberg, S. Wold, QSAR Model Validation BT - Molecular Modeling and Prediction of Bioactivity, in: K. Gundertofte, F.S. Jørgensen (Eds.), Springer US, Boston, MA, 2000: pp. 271–272. https://doi.org/10.1007/978-1-4615-4141-7_36.

[405] S.K. Burley, C. Bhikadiya, C. Bi, S. Bittrich, L. Chen, G. V Crichlow, C.H. Christie, K. Dalenberg, L. Di Costanzo, J.M. Duarte, S. Dutta, Z. Feng, S. Ganesan, D.S. Goodsell, S. Ghosh, R.K. Green, V. Guranović, D. Guzenko, B.P. Hudson, C.L. Lawson, Y. Liang, R. Lowe, H. Namkoong, E. Peisach, I. Persikova, C. Randle, A. Rose, Y. Rose, A. Sali, J. Segura, M. Sekharan, C. Shao, Y.-P. Tao, M. Voigt, J.D. Westbrook, J.Y. Young, C. Zardecki, M. Zhuravleva, RCSB Protein Data Bank: powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences, Nucleic Acids Res. 49 (2021) D437–D451. https://doi.org/10.1093/nar/gkaa1038.

[406] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, UCSF Chimera--a visualization system for exploratory research and analysis., J. Comput. Chem. 25 (2004) 1605–1612. https://doi.org/10.1002/jcc.20084.

[407] O.I. Titov, D.A. Shulga, V.A. Palyulin, N.S. Zefirov, Perspectives of Halogen Bonding Description in Scoring Functions and QSAR/QSPR: Substituent Effects in Aromatic Core, Mol. Inform. 34 (2015) 404–416. https://doi.org/10.1002/minf.201400195.

[408] S. Kortagere, S. Ekins, W.J. Welsh, Halogenated ligands and their interactions with amino acids: implications for structure-activity and structure-toxicity relationships., J. Mol. Graph. Model. 27 (2008) 170–177. https://doi.org/10.1016/j.jmgm.2008.04.001.

[409] V. Khanna, S. Ranganathan, Physiochemical property space distribution among human metabolites, drugs and toxins, BMC Bioinformatics. 10 (2009) 1–18. https://doi.org/10.1186/1471-2105-10-S15-S10.

[410] D.F. Veber, S.R. Johnson, H.Y. Cheng, B.R. Smith, K.W. Ward, K.D. Kopple, Molecular properties that influence the oral bioavailability of drug candidates, J. Med. Chem. 45 (2002) 2615–2623. https://doi.org/10.1021/jm020017n.

[411] E. Estrada, Physicochemical interpretation of molecular connectivity indices, J. Phys. Chem. A. 106 (2002) 9085–9091. https://doi.org/10.1021/jp026238m.

[412] R. Guha, P.C. Jurs, Interpreting computational neural network QSAR Models: A measure of descriptor importance, J. Chem. Inf. Model. 45 (2005) 800–806. https://doi.org/10.1021/ci050022a.

[413] A.N. Jain, Ligand-based structural hypotheses for virtual screening., J. Med. Chem. 47 (2004) 947–961. https://doi.org/10.1021/jm030520f.

[414] W.H.B. Sauer, M.K. Schwarz, Molecular shape diversity of combinatorial libraries: A prerequisite for broad bioactivity, J. Chem. Inf. Comput. Sci. 43 (2003) 987–1003. https://doi.org/10.1021/ci025599w.

[415] P. Torkington, The Moments of Inertia of Molecules with Internal Rotation, J. Chem. Phys. 18 (1950) 407–413. https://doi.org/10.1063/1.1747651.

[416] U. Abdulfatai, A. Uzairu, G.A. Shallangwa, S. Uba, Designing and estimating antioxidant properties of some lubricant additives via QSPR and MD methodologies, Sci. African. 8 (2020) e00451. https://doi.org/https://doi.org/10.1016/j.sciaf.2020.e00451.

[417] E. Byvatov, U. Fechner, J. Sadowski, G. Schneider, Comparison of support vector machine and artificial neural network systems for drug/nondrug classification., J. Chem. Inf. Comput. Sci. 43 (2003) 1882–1889. https://doi.org/10.1021/ci0341161.

[418] M. Siddique, S. Mohanty, D. Panda, A HYBRID FORECASTING MODEL FOR PREDICTION OF STOCK VALUE OF TATA STEEL USING SUPPORT VECTOR REGRESSION AND PARTICLE SWARM OPTIMIZATION, Int. J. Pure Appl. Math. 119 (2018) 1719–1728.

[419] R.D.D. Jawarkar, R.L. Bakal, M.E.A.A. Zaki, S. Al-Hussain, A. Ghosh, A. Gandhi, N. Mukerjee, A. Samad, V.H. Masand, I. Lewaa, QSAR based virtual screening derived identification of a novel hit as a SARS CoV-229E 3CLpro Inhibitor: GA-MLR QSAR modeling supported by molecular Docking, molecular dynamics simulation and MMGBSA calculation approaches, Arab. J. Chem. 15 (2022) 103499. https://doi.org/https://doi.org/10.1016/j.arabjc.2021.103499.

[420] R. Guha, On the interpretation and interpretability of quantitative structure–activity relationship models, J. Comput. Aided. Mol. Des. 22 (2008) 857–871. https://doi.org/10.1007/s10822-008-9240-5.

[421] V. Consonni, R. Todeschini, New Spectral Indices for Molecule Description, MATCH. 60 (2008) 3–14.

[422] G.G. Hall, On the eigenvalues of molecular graphs, Mol. Phys. 33 (1977) 551–557. https://doi.org/10.1080/00268977700100471.

[423] N. Trinajstic, Chemical graph theory, CRC press, 2018.

[424] T. Powell, Bounds on the Ratio of Eigenvalues, (2005).

[425] M. Konwar, D. Sarma, Advances in developing small molecule SARS 3CLpro inhibitors as potential remedy for corona virus infection, Tetrahedron. 77 (2021) 131761. https://doi.org/https://doi.org/10.1016/j.tet.2020.131761.

[426] J. Keeler, P. Wothers, Why chemical reactions happen, Oxford university press, 2003.

[427] C.-C. Chen, X. Yu, C.-J. Kuo, J. Min, S. Chen, L. Ma, K. Liu, R.-T. Guo, Overview of antiviral drug candidates targeting coronaviral 3C-like main proteases., FEBS J. 288 (2021) 5089–5121. https://doi.org/10.1111/febs.15696.

[428] A.D. Rathnayake, J. Zheng, Y. Kim, K.D. Perera, S. Mackin, D.K. Meyerholz, M.M. Kashipathy, K.P. Battaile, S. Lovell, S. Perlman, W.C. Groutas, K.-O. Chang, 3C-like

protease inhibitors block coronavirus replication in vitro and improve survival in MERS-CoV–infected mice, Sci. Transl. Med. 12 (2020) eabc5332. https://doi.org/10.1126/scitranslmed.abc5332.

[429] N. Schneider, R.A. Lewis, N. Fechner, P. Ertl, Chiral Cliffs: Investigating the Influence of Chirality on Binding Affinity, ChemMedChem. 13 (2018) 1315–1324. https://doi.org/https://doi.org/10.1002/cmdc.201700798.

[430] B.S. Karthikeyan, J. Ravichandran, K. Mohanraj, R.P. Vivek-Ananth, A. Samal, A curated knowledgebase on endocrine disrupting chemicals and their biological systems-level perturbations, Sci. Total Environ. 692 (2019) 281–296. https://doi.org/https://doi.org/10.1016/j.scitotenv.2019.07.225.

[431] H.M. Olague, L.H. Etzkorn, S. Gholston, S. Quattlebaum, Empirical Validation of Three Software Metrics Suites to Predict Fault-Proneness of Object-Oriented Classes Developed Using Highly Iterative or Agile Software Development Processes, IEEE Trans. Softw. Eng. 33 (2007) 402–419. https://doi.org/10.1109/TSE.2007.1015.

[432] B. Üstün, W.J. Melssen, L.M.C. Buydens, Visualisation and interpretation of support vector regression models, Anal. Chim. Acta. 595 (2007) 299–309.

[433] H. Okut, Bayesian Regularized Neural Networks for Small n Big p Data, in: J.L.G. Rosa (Ed.), IntechOpen, Rijeka, 2016: p. Ch. 2. https://doi.org/10.5772/63256.

[434] S. Alapour, M.D. Farahani, J.R.A. Silva, C.N. Alves, H.B. Friedrich, D. Ramjugernath, N.A. Koorbanally, Investigation of conventional and non-conventional hydrogen bonds: a comparison of fluorine-substituted and non-fluorine substituted compounds, Monatshefte Für Chemie - Chem. Mon. 148 (2017) 2061–2068. https://doi.org/10.1007/s00706-017-2044-3.

[435] R.C. Wade, P.J. Goodford, The role of hydrogen-bonds in drug binding., Prog. Clin. Biol. Res. 289 (1989) 433–444.

[436] G. Bulusu, G. Desiraju, Strong and Weak Hydrogen Bonds in Protein–Ligand Recognition, J. Indian Inst. Sci. 100 (2019). https://doi.org/10.1007/s41745-019-00141-9.

[437] K.N.-M. Daeffler, H.A. Lester, D.A. Dougherty, Functionally Important Aromatic–Aromatic and Sulfur−π Interactions in the D2 Dopamine Receptor, J. Am. Chem. Soc. 134 (2012) 14890–14896. https://doi.org/10.1021/ja304560x.

[438] Q.-S. Du, Q.-Y. Wang, L.-Q. Du, D. Chen, R.-B. Huang, Theoretical study on the polar hydrogen-π (Hp-π) interactions between protein side chains., Chem. Cent. J. 7 (2013) 92. https://doi.org/10.1186/1752-153X-7-92.

[439] W. Pietruś, R. Kafel, A.J. Bojarski, R. Kurczab, Hydrogen Bonds with Fluorine in Ligand–Protein Complexes-the PDB Analysis and Energy Calculations, Molecules. 27 (2022). https://doi.org/10.3390/molecules27031005.

[440] A. Rana, B. Galmés, A. Frontera, H.S. Biswal, D. Chopra, Unravelling the electronic nature of C–F···O–C non-covalent interaction in proteins and small molecules in the solid state, Phys. Chem. Chem. Phys. 22 (2020) 25704–25711. https://doi.org/10.1039/D0CP05280A.