

**COMPUTATIONAL MODELLING OF THE NUCLEOCAPSID
PROTEIN FOR THE PREDICTION AND DELIVERY OF COVID- 19
DRUG TARGETS**

**BISHUNG, JANET UGUMMAYE
(18PCH01811)**

DECEMBER 2020

**COMPUTATIONAL MODELLING OF THE NUCLEOCAPSID
PROTEIN FOR THE PREDICTION AND DELIVERY OF COVID- 19
DRUG TARGETS**

BY

**BISHUNG, JANET UGUMMAYE
(18PCH01811)**

B.Sc. Computer Science, Landmark University, Kwara State

**A DISSERTATION SUBMITTED TO THE SCHOOL OF
POSTGRADUATE STUDIES IN PARTIAL FULFILMENT OF THE
REQUIREMENTS FOR THE AWARD OF MASTER OF SCIENCE
(M.Sc.) DEGREE IN MANAGEMENT INFORMATION SYSTEM IN
THE DEPARTMENT OF COMPUTER AND INFORMATION
SCIENCES, COLLEGE OF SCIENCE AND TECHNOLOGY,
COVENANT UNIVERSITY.**

DECEMBER 2020

ACCEPTANCE

This is to attest that this dissertation was accepted in partial fulfillment of the requirements for the award of Master of Science (M.Sc.) degree in Management Information System in the Department of Computer and Information Science, College of Science and Technology, Covenant University, Ota, Ogun State, Nigeria.

Mr. John A. Philip

(Secretary, School of Postgraduate Studies)

Signature and Date

Prof. Akan B. Williams

(Dean, School of Postgraduate Studies)

Signature and Date

DECLARATION

I, **BISHUNG JANET UGUMMAYE** with matriculation number **18PCH01811**, hereby declare that this dissertation entitled **COMPUTATIONAL MODELLING OF THE NUCLEOCAPSID PROTEIN FOR THE PREDICTION AND DELIVERY OF COVID-19 DRUG TARGETS** was carried out by me under the supervision of Prof. Victor C. Osamor. This project is an original study in the Department of Computer and Information Sciences, College of Science and Technology, Covenant University, Ota, Nigeria. All scholarly information used in this study is fully acknowledged.

BISHUNG, JANET UGUMMAYE

Signature and Date

CERTIFICATION

This is to certify that the dissertation titled “**COMPUTATIONAL MODELLING OF THE NUCLEOCAPSID PROTEIN FOR THE PREDICTION AND DELIVERY OF COVID-19 DRUG TARGETS**” was carried out by **BISHUNG JANET UGUMMAYE** with matriculation number 18PCH01811 under the supervision of Prof. Victor C. Osamor in the Department of Computer and Information Science, College of Science and Technology, Covenant University, Ota, Ogun State.

Prof. Victor C. Osamor
(Supervisor) **Signature and Date**

Dr. Olufunke O. Oladipupo
(Head of Department) **Signature and Date**

Prof. Olumide B. Longe
(External Examiner) **Signature and Date**

Prof. Akan B. Williams
(Dean, School of Postgraduate Studies) **Signature and Date**

DEDICATION

This dissertation is dedicated to God for his grace and to me for the strength to pull through.

ACKNOWLEDGEMENTS

All praise to God, the creator of the universe, the source of all good inspirations and insights and all wise God for his unlimited grace, mercies, insights, directions, and wisdom all through my academic pursuit in Covenant University. It's His doing and he deserves all the glory.

I would like to appreciate my awesome parents Mr. and Mrs. Matthias Bishung and my siblings for their support physically, spiritually, and financially from the beginning of this academic pursuit to the completion.

Immense appreciation to the Chancellor of Covenant University, Dr David Oyedepo for his vision, which birthed a world class academic institution where sound academic excellence can be obtained, and sound academic research can be made. I also want to appreciate the entire management of Covenant University for seeing that the vision has been fulfilled towards accomplishing great excellence through the master's academic programme.

My sincere gratitude goes to my dissertation supervisor, Prof Victor C. Osamor, for his diligent and great guidance towards the completion of the dissertation. Despite his busy schedule, he made time to guide, instruct and made sure the work was done timely and with excellence. I also appreciate the Head of Department of Computer and Information Sciences (CIS), Dr. Olufunke O. Oladipupo, the PG Coordinator, the former Head of Department, Prof. Ambrose A. Azeta, and the entire faculty members for their remarks during the numerous presentations during the cause of this project work.

Finally, I would like to appreciate my lovely friends Okezie Fiona, Awal Abdulganiyu, Atainyang Idara, Koyejo Ooreofe, Ejiobih Chukwuebuka, Ani Osinachi Sylvester, Edosomwan Boma, Okezie Chidinma, who assisted me in one capacity or the other to make this work a reality, and my colleagues for making this journey a remarkable one despite the challenges we faced throughout the programme.

TABLE OF CONTENT

CONTENT	Page
COVER PAGE	
TITLE PAGE	ii
ACCEPTANCE	iii
DECLARATION	iv
CERTIFICATION	v
DEDICATION	vi
ACKNOWLEDGEMENTS	vii
TABLE OF CONTENT	viii
LIST OF TABLES	xii
LIST OF FIGURES	xiii
LIST OF EQUATIONS	xvi
ABSTRACT	xvii
CHAPTER ONE: INTRODUCTION	1
1.1 Background to the Study	1
1.2 Statement of the Problem	3
1.3 Aim and Objectives of the Study	4
1.4 Research Methodology	5
1.5 Significance of the study	6
1.6 Organization of the Dissertation	6
CHAPTER TWO: LITERATURE REVIEW	7
2.0 Introduction	7
2.1 Coronavirus	7

2.1.1	The Human Coronavirus	12
2.1.2	Life Cycle of COVID-19	13
2.1.3	Testing for COVID-19	15
2.1.4	Genomic Sequencing	17
2.2	Artificial Intelligence Tools	18
2.3	Computational Modelling	21
2.3.1	ODE-Based Modeling	22
2.3.2	Petri Net-Based Modeling Approaches	24
2.3.3	Boolean Modeling Approaches	25
2.3.4	Agent-Based Model of Biological System	25
2.3.5	Rule-based Modelling	26
2.3.6	Linear Programming Approaches	27
2.4	Related Techniques	27
2.4.1	Machine Learning	28
2.4.1.1	Types of Machine Learning Techniques	28
2.4.1.2	Types of Machine Learning Approach	29
2.4.1.3	Training of Neural Networks	30
2.4.2	Drug Discovery	32
2.4.3	COVID-19 Ongoing Clinical Trials	35
2.5	Review of Related Works	36
CHAPTER THREE: METHODOLOGY		40
3.1	Introduction	40
3.2	Methodology	41
3.2.1	Acquisition of Genomic Sequence data (objective1)	41
3.2.2	Data Analysis (objective2)	43

3.2.2.1	Phylogeny	43
3.2.2.2	Motif Discovery	45
3.2.3	Homology Modelling (objective 3)	47
3.2.4	Druggable remedy prediction (objective 4)	48
CHAPTER FOUR: RESULT AND DISCUSSION		50
4.1	Introduction	50
4.2	Implementation Tools/Techniques	50
4.3	Data Acquisition	51
4.4	Data Analysis	52
4.4.1	Phylogeny	53
4.4.1.1	Sequence Alignment	53
4.4.1.2	ORF Region Finding	57
4.4.1.3	Phylogeny Analysis	60
4.4.2	Motif Discovery	62
4.4.2.1	Preprocessing	63
4.4.2.2	Discovering	65
4.4.2.3	Post Processing	66
4.5	Homology Modelling	68
4.5.1	Identification and Selection of Templates	69
4.5.2	Sequence Alignment and Alignment Corrections	70
4.5.3	Model Building	72
4.5.4	Model Validation	75
4.6	Druggable Remedies	77
4.6.1	Initializing Molecules	77
4.6.2	Running AutoGrid	81

4.6.3	Running AutoDock Vina	83
4.6.4	Analyzing Interaction Energy	88
4.7	Discussion	89
CHAPTER FIVE: CONCLUSION AND RECOMMENDATION		93
5.1	Summary	93
5.2	Conclusion	93
5.3	Contributions to Knowledge	94
5.4	Recommendation	94
REFERENCES		95
APPENDIX		107

LIST OF TABLES

Table	Title of Table	Page
1.1:	Objective and Methodology Mapping	6

LIST OF FIGURES

Figure	Title of Figures	Page
1.1:	Countries, territories, or areas with reported confirmed cases of COVID-19	2
1.2:	A graphical illustration of the rapid increase in the confirmed cases	2
2.1:	A graphical illustration of the mode of transmission of the Coronaviridae family	8
2.2:	3D illustration of coronavirus	9
2.3:	Coronavirus illustration	10
2.4:	The life cycle of COVID-19	14
2.5:	Transmission mode of COVID-19	14
2.6:	A Pictorial view of how the nasopharyngeal swabs are collected from patients	16
2.7:	A DNA molecule	17
2.8:	Pictorial Illustration of a Whole Genome Sequencing	18
2.9:	AI companies and Biotech firm's partnership	20
2.10:	A simple neuron description.	30
2.11:	The application of AI at each stage of drug development	34
3.1:	Process flow depicting the methodology for COVID-19 drug discovery.	40
3.2:	Chain sequence data of the SARS-CoV Spike protein	41
3.3:	3D structural view of the SARS-CoV Spike protein	42
3.4:	Sample of the SARS-CoV2 nucleocapsid protein sequence chain.	42
3.5:	The phylogenetic analysis procedure.	45
3.6:	Motif discovery steps	46
3.7:	Steps in homology modelling	47
4.1:	Sequence data set of the SARS-CoV- 2/human/AUS/VIC06/2020	52
4.2:	Representation of the ClustalW neighbour joining algorithm	54
4.3:	Screenshot of the ClustalW Alignment settings in MEGAX	55
4.4:	Screenshot of the ClustalW matrix settings in MEGAX	55
4.5:	Screenshot of the merged sequences prior to alignment.	57
4.6:	Screenshot of the merged sequences after alignment.	57
4.7:	Screenshot of ORF finder platform	58
4.8:	Screenshot of the ORF findings and the protein translation of ORF29	59

4.9: A screenshot of the ORF findings and the protein translation of ORF51	59
4.10 A screenshot of the ORF findings and the protein translation of ORF51	60
4.11: Screenshot of the dialog progress box of the phylogeny tree creation.	61
4.12: Evolutionary analysis using the Maximum Likelihood method	62
4.13: Unprocessed RNA sequence	64
4.14: Processed RNA sequence.	65
4.15: A screenshot of the discovery process on MEME suite.	66
4.16: Screenshot of the discovered motif.	67
4.17: screenshot motif summary.	68
4.18 :Screenshot of the blast result for a likely template.	69
4.19: Demonstrating a perspective for the nucleocapsid proteins for the modeling.	70
4.20: A screenshot of the python script used in the alignment process.	71
4.21: A screenshot of a cross-section of the aligned sequences.	71
4.22: Source code for the building of the model.	72
4.23: A screenshot of the modeller being used to build the model structure	73
4.24: Screenshot of the first generated model structure of the nucleocapsid protein	74
4.25 :The second generated model structure of the nucleocapsid protein	74
4.26: Validated modelled structure for SARS-CoV2 Nucleocapsid Protein	75
4.27: Ramachandran plot showing 95.9% success rate in the structured model	76
4.28: Screenshot of the N Protein after the removal of water molecules.	78
4.29: Screenshot of the NProtein after the addition of polar hydrogen molecules	80
4.30: Screenshot of adding Kolman charges to the protein. 7.0 charges were added	80
4.31: Merging processes involved in the docking process.	81
4.32: Graphical illustration of the blind docking process.	82
4.33: Saved text and pdbqt formats of the N protein and hydroxychloroquine sulfate	84
4.34: Saved text and pdbqt formats of the N protein and diallyl thiosulfate	85
4.35: Saved text and pdbqt formats of the N protein and Isotretinoin	85
4.36: Ribbon like docking result structures	86
4.37 Surface like docking result structures	87
4.38 The interacting amino acids within the Diallyl thiosulfate ligand.	88
4.39 The interacting amino acids within the hydroxychloroquine sulfate ligand.	88

4.40: The interacting amino acids within the Isotretinoin ligand.

89

LIST OF EQUATIONS

Equation	Title of Equation	Page
(1)	Law of Mass Action	22
(2)	Comprehensive Law of Mass Action	22
(3)	Hill Function	23
(4)	Michaelis-Menten Kinetics	23
(5)	Maximum Likelihood	61

ABSTRACT

The Severe acute respiratory syndrome coronavirus 2, which is the causative agent of the Coronavirus Disease 2019 (COVID-19) pandemic is more than a threat to public health. This is due to an inadequate understanding of the molecular keys that constitute its viral protein chemistry. Since its outbreak in December 2019, there have been a series of research, trials, tests carried out and genomic analysis was done to ascertain infection, hinder transmission, and evolve clinical intervention. With several deaths and no hope for a cure anytime soon, more in-depth knowledge of the virus is urgently needed. This project aims to investigate the structural, molecular makeup, and epidemiology of Coronavirus and further make use of computational models to derive a three-dimensional structure of the Nucleocapsid protein of the virus in order to find drug targets that can help inhibit the replication of the virus and make it visible for the human immune system to fight off the virus. During this work, the structural, molecular makeup, and epidemiology of Coronavirus were analyzed alongside a detailed understanding of the selected nucleocapsid protein. Homology modeling and model validation was performed to confirm the authenticity of the structure predicted. The result showed a likelihood of the N-protein being a suitable protein to be used in its test phase because of its stability and the various drug targets found during the work. With the knowledge harnessed, we will be a step closer to finding a therapeutic drug that will inhibit the spread of this deadly virus.

Keywords: Drug targets, COVID-19, Nucleocapsid Protein, protein, clinical trials. Artificial intelligence