



UNIVERSITY OF  
KWAZULU-NATAL

---

INYUVESI  
YAKWAZULU-NATALI

***In Silico* Investigation of Hepatitis C Virus: A Novel  
Perspective into Targeted Viral Inhibition of NS3  
Helicase, NS3/4A Protease and NS5B RNA-  
Dependent RNA Polymerase**

Miss Letitia Shunmugam

2019

A thesis submitted to the School of Health Sciences, University of KwaZulu-Natal,  
Westville, in fulfilment for the degree of Doctor of Philosophy

***In Silico* Investigation of Hepatitis C Virus: A Novel Perspective into Targeted Viral  
Inhibition of NS3 Helicase, NS3/4A Protease and NS5B RNA-Dependent RNA  
Polymerase**

**Miss Letitia Shunmugam**

**211503447**

**2019**

A thesis submitted to the School of Pharmacy and Pharmacology, Faculty of Health Sciences,  
University of KwaZulu-Natal, Westville, for the degree of Doctor of Philosophy.

This is the thesis in which the chapters are written as a set of discrete research publications, with an  
overall introduction and final summary. Typically, these chapters will have been published in  
internationally recognized, peer-reviewed journals.

This is to certify that the contents of this thesis are the original research work of Miss Letitia  
Shunmugam.

As the candidate's supervisor, I have approved this thesis for submission.

Supervisor:

Signed:

Name: Prof Mahmoud E.S. Soliman

Date:

## PREFACE

This thesis is divided into eight chapters, including this one:

### **Chapter 1:**

This is an introductory chapter that addresses the background, rationale and relevance of the study as well as the proposed aim and objectives. The general outline and structure of the thesis concludes this chapter.

### **Chapter 2:**

This chapter provides a comprehensive literature review on HCV and the urgent need for continuous research that will facilitate the design, discovery and development of more effective and potent potential inhibitors, which address challenges that hinder the efficacy of current market-available anti-HCV drugs. This chapter includes the discovery, epidemiology, transmission, life cycle, viral diagnostics, pathogenesis, clinical features, viral characteristics (mechanistic and structural), viral/host drug targets, specifically the NS5 and NS3 protein and explores the field of covalent inhibition.

### **Chapter 3:**

This chapter explains computer aided drug design by discussing various molecular dynamic approaches and applications. The computational tools and software required to study the comparative inhibitory modes of proteins as well as methods to analyse binding affinity are also explained. This chapter conceptualizes computer-aided drug design by discussing various molecular modelling and molecular dynamic techniques and applications. The computational tools needed to investigate comparative enzymatic structural/conformational characteristics as well as methods used to analyse binding affinity is elucidated upon.

### **Chapter 4: (Published work- this chapter is presented in the required format of the journal and is the final version of the accepted manuscript)**

This chapter provides an all exclusive “road map” entitled “Road Map for the Structure-Based Design of Selective Covalent HCV NS3/4A Protease Inhibitors”. This study provides a comprehensive overview from literature and presents technical guidance that may be used to initiate a systematic “road map” for the design of selective covalent inhibitors that may assist medicinal chemists and scientists in the design and development of optimized potential selective covalent HCV NS3/4A viral protease inhibitors. This article was published in The Protein Journal (impact factor (IF) = 1.133).

**Chapter 5: (Published work- this chapter is presented in the required format of the journal and is the final version of the accepted manuscript)**

This chapter investigates the second objective of the thesis and is titled “Targeting HCV Polymerase: A Structural and Dynamic Perspective into the Mechanism of Selective Covalent Inhibition”. This study implements an optimised molecular modelling approach to better understand the structural and dynamic impact of selective covalent inhibitor binding to HCV NS5B RNA-dependent RNA polymerase (RdRp). This article has been published in RSC Advances (IF = 3.049).

**Chapter 6: (This chapter is presented in the required format of the journal and is the final version of the submitted manuscript)**

This chapter, “Enhanced Pharmacophore-based Virtual Screening Approach in the Discovery of Potential Inhibitors of HCV NS3 Helicase”, assesses the third objective of the thesis: To identify potential HCV NS3 helicase inhibitors through utilisation of the pharmacophore-based virtual screening approach using the chemical scaffold of quercetin as a model. This study implements an enhanced and proven screening approach to identify lead compounds that may be potentially used as inhibitors of HCV NS3 helicase. This computational “per-residue energy decomposition pharmacophore” virtual screening approach will be impetus in initiating the discovery of effective inhibitors of HCV viral targets. This article has been submitted.

**Chapter 7:**

This is the final chapter that proposes future work and concluding remarks.

## ABSTRACT

Hepatitis C Virus (HCV) is an escalating global healthcare and economic burden that requires extensive intervention to alleviate its control. Over the years, drug design efforts have produced many anti-HCV drugs; however, due to drug resistance brought on by numerous genetic variations of the virus and lack of specificity and stability, current drugs are rendered ineffective. The situation has been further intensified by the absence of a viable vaccine. For these reasons, continuous HCV research is imperative for the design and development of promising inhibitors that address the challenges faced by present antiviral therapies. Moreover, exposure of previously neglected viral protein targets can offer another potentially valuable therapeutic route in drug design research.

Structure-based drug design approaches accentuate the development of small inhibitor molecules that interact with therapeutic targets through non-covalent interactions. The unexpected discovery of covalent inhibitors and their distinctive nature of instigating complete and irreversible inhibition of targets have shifted attention away from the use of non-covalent drugs in antiviral treatment. This has led to significant progress in understanding covalent inhibition regarding their underlying mechanism of action and in the design of novel covalent inhibitors that work against biological targets. However, due to difficulties arising in its application and resultant safety, the pharmaceutical industry were reluctant to pursue this strategy. With the use of rational drug design, a novel strategy was then proposed known as selective covalent inhibition. Due to the lack of competent protocols and information, little is known regarding selective covalent inhibition

This study investigates three biological HCV targets, NS3 protease, RNA helicase and NS5B RNA-dependent RNA polymerase. With constantly evolving viruses like HCV, computational methods including molecular modelling and docking, virtual screening and molecular dynamic simulations have allowed chemists to screen millions of compounds to filter out potential lead drugs. These *in silico* approaches have allowed Computer-Aided Drug Design as a cost-effective strategy to accelerate the process of drug discovery.

The above techniques, with numerous other computational tools were employed in this study to fill the gap in HCV drug research by providing insights into the structural and dynamic changes that describe the mechanism of selective covalent inhibition and pharmacophoric features that lead to unearthing of potential small inhibitor molecules against Hepatitis C.

The first study (Chapter 4) provides a comprehensive review on HCV NS3/4A protein, current therapies and covalent inhibition as well as introduces a technical guideline that provides a systematic approach for the design and development of potent, selective HCV inhibitors.

The second study (Chapter 5) provides a comprehensive understanding concerning the implications of selective covalent inhibition on the activity of HCV NS5B RNA-dependent RNA polymerase, with respect to key components required for viral replication, when bound to a target-specific small inhibitor molecule.

The third study (Chapter 6) is preliminary investigation that uses Pharmacophore-based virtual screening as an efficient tool for the discovery of improved potential HCV NS3 helicase inhibitors. The pharmacophoric features were created based on the highly contributing amino acid residues that bind with highest affinity to the weak inhibitor, quercetin. These residues were identified based on free energy footprints obtained from molecular dynamic and thermodynamic calculations. Post molecular dynamic analysis and appropriate drug-likeness properties of the three top-hit compounds revealed that ZINC02495613 could be a more effective potential HCV helicase inhibitor; however, further validation steps are still required.

This study offers a comprehensive *in silico* perspective to fill the gap in rational drug design research against HCV, thus providing an insight into the mechanism of selective covalent inhibition, uncovering a previously neglected viral target and identifying possible antiviral drugs. To this end, the work presented in this report is considered a fundamental platform to advance research toward the design and development of novel and selective anti-HCV drugs.

## DECLARATION I – PLAGIARISM

I, Letitia Shunmugam, declare that

1. The research reported in this thesis, except where otherwise indicated, is my original research.
2. This thesis has not been submitted for any degree or examination at any other university.
3. This thesis does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.
4. This thesis does not contain other person's writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:
  - a. Their words have been re-written, but the general information attributed to them has been referenced.
  - b. Where their exact words have been used, then their writing has been placed in italics and inside quotation marks and referenced.
5. This thesis does not contain text, graphics or tables copied and pasted from the internet, unless specifically acknowledged, and the source being detailed in the thesis and in the reference sections.

A detailed contribution to publications that form part or/and include research presented in this thesis is stated (include publications submitted, accepted, in press and published).

**Signed: L. Shunmugam**

## DECLARATION II – LIST OF PUBLICATIONS

1. Letitia Shunmugam, Pritika Ramharack and Mahmoud E.S. Soliman (2017). Road Map for the Structure-Based Design of Selective Covalent HCV NS3/4A Protease Inhibitors. *The Protein Journal*, 36(5), 397-406. (Published)

### **Contribution:**

Letitia Shunmugam: contributed to the project by performing all the experimental work and manuscript preparation and writing.

Pritika Ramharack: contributed to the project by assisting in manuscript writing.

Mahmoud E.S. Soliman: Supervisor

*Appendix B: PDF version of the publication*

2. Letitia Shunmugam and Mahmoud E.S. Soliman (2018). Targeting HCV Polymerase: A Structural and Dynamic Perspective into the Mechanism of Selective Covalent Inhibition. *RSC Advances*, 8(73), 42210-42222 (Published)

### **Contribution:**

Letitia Shunmugam: contributed to the project by performing all the experimental work and manuscript preparation and writing.

Mahmoud E.S. Soliman: Supervisor

*Appendix C: PDF version of the publication*

3. Letitia Shunmugam, Nikita Devnarain and Mahmoud E.S. Soliman (2018). Enhanced Pharmacophore-based Virtual Screening Approach in the Discovery of Potential Inhibitors of HCV NS3 Helicase, *Future Virology*, Manuscript ID: FVL-2019-0006 (Submitted)

### **Contribution:**

Letitia Shunmugam: contributed to the project by performing all the experimental work and manuscript preparation and writing.

Nikita Devnarain: contributed to the project by assisting in experimental work and manuscript writing.

Mahmoud E.S. Soliman: Supervisor

*Appendix D: PDF version of the submission*



## RESEARCH OUTPUT

### A- LIST OF PUBLICATIONS

1. Letitia Shunmugam, Pritika Ramharack and Mahmoud E.S. Soliman (2017). Road Map for the Structure-Based Design of Selective Covalent HCV NS3/4A Protease Inhibitors. *The Protein Journal*, 36(5), 397-406.
2. Letitia Shunmugam and Mahmoud E.S. Soliman (2018). Targeting HCV Polymerase: A Structural and Dynamic Perspective into the Mechanism of Selective Covalent Inhibition. *RSC Advances*, 8(73), 42210-42222.

### B- SUBMITTED

3. Letitia Shunmugam, Nikita Devnarain and Mahmoud E.S. Soliman (2018). Enhanced Pharmacophore-based Virtual Screening Approach in the Discovery of Potential Inhibitors of HCV NS3 Helicase, *Future Virology*, Manuscript ID: FVL-2019-0006 (*Submitted*)

### C- CONTRIBUTION AS CO-AUTHOR

4. Emmanuel Adefemi Adeniji, Fisayo Olutu, Letitia Shunmugam and Mahmoud E.S. Soliman. Covalent inhibition of human liver mitochondrial aldehyde dehydrogenase-2 by aldehydic products of lipid peroxidation disrupts co-enzyme binding and oligomerization: A dynamical and structural perspective, *Journal of Molecular Simulation*, Manuscript ID: GMOS-2018-0301.R1 (*Accepted*)
5. Pritika Ramharack, Nikita Devnarain, Letitia Shunmugam and Mahmoud E.S. Soliman. Navigating Research Toward the Re-emerging Nipah Virus- A New Piece to the Puzzle, *Expert Opinion on Drug Discovery*, Manuscript ID: EODC-2018-ST-0130 (*Submitted*)

6. Nikita Devnarain, Letitia Shunmugam and Mahmoud E.S. Soliman. Chronicles of viperin: Are we using the most of this broad-spectrum human weapon to counteract viruses? *Drug Discovery Today*, Manuscript ID: DRUDIS\_2018\_268 (*Submitted*)
  
7. Abdolkarim Farrokhzadeh, Farideh Badichi Akher, Pritika Ramharack, Letitia Shunmugam and Mahmoud E.S. Soliman. Discovery of Novel Natural Flavonoids as Potent Antiviral Candidates against Hepatitis C Virus NS5B Polymerase, *Future Virology*, Manuscript ID: FVL-2019-0007 (*Submitted*)

## **DEDICATION**

To my late father, Cyril Shunmugam.

(07/01/1960-30/05/2017)

I am eternally grateful for all the sacrifices you have made as I've grown up. I would not be the person I am today if it had not been for your encouragement, support and unconditional love. I'm so blessed and honoured to have been raised by a man like you. The legacy you left behind will forever remain in my heart and of those around us. Thank you, Dad, you will always be my hero. I love you.

## ACKNOWLEDGEMENTS

Firstly, all praises and thanks be to God, without his assistance and grace, nothing can be possible.

I would also like to thank:

**My family**, for the love, support and patience without which the completion of this thesis would not have been possible. To my Mum, your prayers have seen me through the darkest of days, you have cared for me and loved me especially through this difficult and demanding time. No one can ever replace you in my life. To my brother, thank you for being my shoulder to cry on. These past two years have not been easy but knowing you are by my side gives me comfort. To my Dad, I miss you dearly, I know that you are guiding and protecting me. I am truly grateful for the morals and values you have instilled within me, I hope I make you proud.

**Professor Mahmoud E.S. Soliman**, I have the utmost gratitude for the motivation, constructive criticism and encouragement which you have provided throughout my Doctoral studies.

**The Molecular Bio-computation and Drug Design Laboratory**, Thank you all for the friendship, support and assistance. The memories created shall forever be a reminder of the friendships made.

**National Research Foundation (DAAD) and UKZN College of Health Science**, for the scholarship that provided financial assistance for completion of this degree.

**The Centre of High-Performance Computing (CHPC)** for providing computational resources.

## LIST OF ABBREVIATIONS

$\alpha$	Alpha
$\theta$	Angstrom
$\text{\AA}$	Angle
$\beta$	Beta
$^{\circ}$	Degree
$\Delta$	Delta
%	Percentage
3D	Three-dimensional
BFE/ $\Delta G$	Binding free energy
CADD	Computer-aided drug design
CMD	Covalent molecular dynamics
DCCM	Dynamic cross correlation matrix
GAFF	General Amber force field
GPU	Graphics Processing Unit
HCV	Hepatitis C Virus
K	Kelvin
kDa	Kilodalton
LBVS	Ligand-based virtual screening
MC	Monte Carlo
MD	Molecular dynamics
MM	Molecular Mechanics
MM/GBSA	Molecular Mechanics/Generalized Born Surface Area

MMV	Molecule molecular viewer
mol	Mole
NI	Nucleoside/nucleotide inhibitors
NNI	Non-nucleoside inhibitors
NMR	Nuclear magnetic resonance
NPT	Isobaric–isothermal ensemble
ns	Nanoseconds
NS5B	Non-structural protein 5B
PBVS	Pharmacophore-based virtual screening
PDB	Protein data Bank
PMEMD	Particle Mesh Ewald Molecular Dynamic
ps	Picosecond
QM	Quantum Mechanics
RDD	Rational drug design
RdRp	RNA-dependent RNA polymerase
RESP	Restrained electrostatic potential
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation
RNA	Ribonucleic Acid
RoG	Radius of gyration
RSCB	Research Collaboratory for Structural Bioinformatics
SASA	Solvent accessible surface analysis

SBVS	Structure-based virtual screening
VMD	Visual molecular dynamics
VS	Virtual screening
WHO	World health organisation

## LIST OF AMINO ACIDS

<b>Three Letter Code</b>	<b>Amino Acid</b>
Ala	Alanine
Arg	Arginine
Asn	Asparagine
Asp	Aspartic Acid
Cys	Cysteine
Gln	Glutamine
Glu	Glutamic Acid
Gly	Glycine
His	Histadine
Ile	Isoleucine
Leu	Leucine
Lys	Lysine
Met	Methionine
Phe	Phenylalanine
Pro	Proline
Ser	Serine
Thr	Threonine
Trp	Tryptophan
Tyr	Tyrosine
Val	Valine



## LIST OF FIGURES

### CHAPTER 2

- Figure 1:** Geographical representation of global HCV genotype distribution (Adapted from Gower *et al.*, 2014). 10
- Figure 2:** Risk factors associated with transmission of HCV (Adapted from Stambouli, 2014). 11
- Figure 3:** Schematic representation of hepatitis C virus (HCV) genome organisation of the polyprotein and viral particle (Prepared by author). 12
- Figure 4:** Overview of HCV life cycle in liver cells (Adapted from Wang *et al.*, 2017). 15
- Figure 5:** Overview of HCV pathogenesis (Prepared by Author). 20
- Figure 6:** The three-dimensional crystal structure of HCV NS3 protein complex (PDB ID: 4B6E) highlighting important elements: the RNA helicase (green), protease (blue), NS4A (magenta), Arg-clamp and Phe-loop (purple), catalytic triad and the ATP-binding domain, using UCSF Chimera software program (Prepared by Author). 21
- Figure 7:** The three-dimensional crystal structure of HCV NS5B polymerase (PDB ID: 3H5S) highlighting the three domains and important elements required for viral replication, generated using UCSF Chimera software program (Prepared by Author). 23
- Figure 8:** Schematic representation of conventional non-covalent (A) and covalent (B) drug interactions (Adapted from Bauer, 2015). In the first step (A) The drug binds to the target protein through a conventional non-covalent reversible inhibitory mechanism, which results in the formation of a protein-drug complex. (B) A non-covalent interaction initially occurs then a subsequent chemical reaction takes place between the protein and drug, in doing so, a covalent irreversibly bound complex is generated (Bauer, 2015; Shunmugam, Ramharack and Soliman, 2017). 31
- Figure 9:** Percentage distribution of covalent drugs currently used in the pharmaceutical industry (Guterman, 2011). 32
- Figure 10:** List of FDA approved covalent inhibitors (Kumalo, Bhakat and Soliman, 2015). 33

### CHAPTER 3

- Figure 1:** Schematic representation of (A) Niels Bohr and (B) Erwin Schrödinger atom models (Prepared by Author). 51

**Figure 2:** Two-dimensional graphical representation of potential energy surface (Adapted from The California State University 2017). 54

**Figure 3:** Diagrammatic illustration of the total potential energy function of a molecule (Adapted from Saenz-Mendez *et al.*,2017) 56

## CHAPTER 4

**Figure 1:** The schematic representation of HCV genomic polyprotein organisation. The 3000 kDa 5'-3' untranslated region (UTR) is comprised of 4 conserved domains which translate into 4 structural proteins: core (C), envelope proteins 1 and 2 (E1/E2) and hydrophilic ion channel protein 7 (P7). The NS3/4A protease mediates the cleavage of the HCV polyprotein resulting in the generation of non-structural proteins: NS2, NS4A co-factor, interferon resisting NS5A protein and NS5B RNA dependent RNA polymerase 4. 77

**Figure 2:** (A) The crystal structure of NS3/4A serine protease holoenzyme. (PDB ID: 4B6E). (B) Representation of the protease domain of NS3 in complex with NS4A (PDB ID: 4U01). The catalytic triad residues of the active site are positioned in the cleft which connects the helicase and protease domains. The catalytic binding site is shallow as indicated in the insert. Zinc has been identified as an important structural feature however its exact function is unknown. Protease domain: blue; helicase domain: green; NS4A co-factor: magenta; catalytic triad: orange; active binding site: yellow; zinc: black. 80

**Figure 3:** Examples of non-covalent and covalent inhibitors including their HCV NS3/4A protein target (PDB: 2OC8). All non-covalent and covalent chemical structures were obtained using PubChem. Non-covalent inhibitors generally dock to the shallow binding pocket of NS3/4A protease and held in place by intermolecular forces. Contrastingly, covalent inhibitors specifically target and bind to the S139 residue located at the catalytic site of the HCV protease to initiate inhibition. The red arrows represent the inhibitor atom that undergoes covalent modification to facilitate the formation of a covalent bond between the inhibitor and S139. NS3 protease domain: blue; NS4A co-factor: magenta; active binding site: yellow; S139: red. 83

**Figure 4:** Alignment of NS3 viral protease sequence from 4 members of the *Flavivirus* family. S139 is conserved within most proteases however C159 is only present in the HCV NS3 protease sequence. DV: Dengue virus; WNV: West Nile virus; S: serine 139 (red); C: cysteine 159 (blue); \* indicates identical residues. 85

**Figure 5:** Sequence alignment comparing HCV viral protease with Homo sapien derived cathepsin K protease. Cathepsin K is 1 of 500 proteases found in Homo sapiens and is functional in bone resorption. As depicted, there are some structural 86

similarities (green) however cathepsin K lacks S139 (red) and C159 (blue). \* indicates identical residues.

**Figure 6:** Summarised workflow adopted *in silico* by selective covalent inhibitor drug design. 90

## CHAPTER 5

**Figure 1:** Graphical representation of right-hand x-ray crystallography structure of NS5B RNA dependent RNA polymerase (PDB ID: 3H5S). The active site is denoted by \*.

**Figure 2:** Sequence alignment of NS5B polymerase. Cysteine 366 is conserved within all genotypes of HCV. C: cysteine (red); Yellow star indicates identical residues. The genome sequences for each of the HCV genotypes were obtained from UniProt and subsequently aligned using ClustalW2<sup>27,28</sup>.

**Figure 3:** Two-dimensional structural representation of HCV RdRp covalent inhibitor, compound 47. Red: C-3 pyridone ring; Blue: indole core and Yellow: benzene ring. The compound exhibited covalent binding to RdRp whereby the thiol group of Cys366 attacked the benzene ring at the position para to the nitro group and the fluoro group is released presumably through an aromatic nucleophilic substitution reaction  $S_NAr$ .

**Figure 4:** RMSD plot of C- $\alpha$  atoms of the apo and bound systems. 110

**Figure 5:** Residue-based average C- $\alpha$  fluctuations of the apo and bound conformation of HCV RdRp throughout 200ns MD simulation. 111

**Figure 6:** Distance analysis demonstrating the conformational changes of the TCW and interdomain angle upon compound 47 binding. **(A)** Interdomain angle was computed by measuring the angle between the centre of masses of the finger, thumb and palm domains. **(B)** The width is calculated by measuring the distance between the C- $\alpha$  atoms of residues Met139 and Val405. **(C)** Structural dynamic movements of the TCW in the apo (purple) and bound (orange, ligand highlighted in green) states of HCV RdRp throughout a 200ns molecular dynamics simulation. The dashed line in the highlighted images represents the distance between the C- $\alpha$  atoms of the two residues. 114

**Figure 7:** Dynamical cross-correlation matrix presenting correlation of residues in the apo **(A)** and compound 47-bound system **(B)**. The status of correlated motions is deduced by the colour scale on the right, the black cross indicates Cys366 to which compound 47 is covalently bound. 115

**Figure 8:** (A) Solvent accessible surface area of apo and compound 47 bound RdRp. (B) Radius of gyration of both systems were measured over a 200ns simulation. Data obtained was represented as a plot displaying the differences arising in radius deviation between the apo and covalent systems. 116

**Figure 9:** Ligand-residue interaction diagram of compound, 47 inside the active site of the palm domain of the RdRp. Orange dashed lines denote hydrogen bond interactions. 117

**Figure 10:** The per-residue energy decomposition analysis of compound 47 bound HCV RdRp. 119

## CHAPTER 6

**Figure 1:** Three-dimensional structure of NS3 protein complex (PDB: 3O8D). The ATP-binding domain (highlight in red) is the region in which ATP or Quercetin can effectively bind. 129

**Figure 2:** The systematic workflow applied in this study. 130

**Figure 3:** Pharmacophore model generation from quercetin-helicase complex. The highlighted yellow regions represent the pharmacophoric moieties that were selected for the model, based on residues that contributed the most to the binding affinity. Using ZINCPharmer, three hits were generated from the pharmacophore-based virtual screening process. 134

**Figure 4:** (A) The per-residue energy decomposition analysis and (B) ligand interaction plot of ZINC18022381 bound to the ATP-binding site of HCV NS3 helicase. 137

**Figure 5:** (A) The per-residue energy decomposition analysis and (B) ligand interaction plot of ZINC33753131 bound to the ATP-binding site of HCV NS3 helicase. 138

**Figure 6:** (A) The per-residue energy decomposition analysis and (B) ligand interaction plot ZINC02495613 bound to the ATP-binding site of HCV NS3 helicase. 139

## LIST OF TABLES

### CHAPTER 2

<b>Table 1:</b> Current HCV direct-acting antiviral drugs (DAAs) in Clinical Testing or FDA approved.	25
---	----

### CHAPTER 4

<b>Table 1:</b> <i>In silico</i> software available for SBDD of selective covalent HCV viral protease inhibitors.	91
---	----

### CHAPTER 5

<b>Table 1:</b> Summary of MM/GBSA-based binding free energy contributions to the compound 47- HCV RdRp complex	118
---	-----

### CHAPTER 6

<b>Table 1:</b> The ID codes, docked 3D-structures and Lipinski's rule of five for the three hits obtained through virtual screening using ZINCPharmer [49,56,57].	135
--	-----

<b>Table 2:</b> Summary of MM/GBSA-based binding free energy contributions of Quercetin and the top-ranked compounds with HCV NS3 helicase.	136
---	-----

## TABLE OF CONTENTS

PREFACE.....	ii
ABSTRACT .....	iv
DECLARATION I – PLAGIARISM .....	vi
DECLARATION II – LIST OF PUBLICATIONS .....	vii
RESEARCH OUTPUT.....	viii
DEDICATION.....	x
ACKNOWLEDGEMENTS.....	xi
LIST OF ABBREVIATIONS.....	xii
LIST OF FIGURES .....	xvi
LIST OF TABLES.....	xxi
CHAPTER 1 .....	1
1. Introduction.....	1
1.1 Background and Rationale of the Study .....	1
1.2 Aims and Objectives.....	3
1.3 Novelty and Significance of the Study .....	4
CHAPTER 2 .....	8
2. Background on the Hepatitis C Virus .....	8
2.1 Introduction.....	8
2.2 Discovery and Epidemiology .....	8
2.3 Route of Transmission.....	10
2.4 Molecular Virology .....	11
2.4.1 Genome Organisation .....	11
2.4.2 Viral Proteins .....	12
2.4.2.1 Structural.....	12
2.4.2.2 Non-structural.....	13
2.4.3 Life cycle of HCV .....	14
2.5 Clinical Features .....	15
2.5.1 Acute HCV Infection.....	15

2.5.2 Chronic HCV Infection.....	16
2.6 Pathogenesis .....	16
2.7 HCV NS3.....	20
2.7.1 Structure, Function and Potential as a Drug Target.....	20
2.8 HCV NS5B .....	22
2.8.1 Structure, Function and Potential as a Drug Target.....	22
2.9 HCV therapy .....	24
2.9.1 Current Therapies .....	24
2.9.1.1 Ribavirin and Pegylated Interferon-alpha.....	24
2.9.1.2 Direct-Acting Antiviral Drugs.....	25
2.9.2 Drug Resistance .....	30
2.10 Covalent Inhibition .....	31
2.10.1 Selective Covalent Inhibition .....	34
CHAPTER 3 .....	48
3. Molecular Modelling and Computational Approaches to Biomolecular Structure and Covalent Drug Design.....	48
3.1 Introduction.....	48
3.2 The principle of Quantum Mechanics .....	50
3.2.1 The Schrödinger Wave Function .....	50
3.2.2 The Born-Oppenheimer Approximation Theory .....	52
3.2.3 Potential Energy Surface as an Application of Quantum Mechanics.....	53
3.3 The Principle of Molecular Mechanics.....	54
3.3.1 Potential Energy Function .....	55
3.4 The Hybrid Quantum Mechanics/Molecular Mechanics.....	57
3.5 The Principle of Molecular Dynamic Simulations .....	58
3.5.1 Covalent Molecular Dynamic Simulation .....	59
3.5.2 Molecular Dynamics Post-Analysis .....	60
3.5.2.1 System Stability .....	61
3.5.2.2 Thermodynamic Energy Calculations (Free Binding Energy) .....	62



3.5.2.3 Conformational Features of System .....	63
3.5.2.4 Dynamic Cross Correlation Matrix of a System (DCCM): .....	64
3.6 Other Computer-Aided Drug Design Techniques Utilized in the Study .....	64
3.6.1 Covalent Docking .....	64
3.6.2 Molecular Docking .....	65
3.6.3 Virtual Screening (VS) .....	65
CHAPTER 4 .....	75
Road Map for the Structure-Based Design of Selective Covalent HCV NS3/4A Viral Protease Inhibitors .....	75
Introduction.....	77
HCV protease as a critical target in drug discovery .....	78
Non-covalent vs covalent inhibition of HCV NS3 protease.....	81
Exploring the concept of “selective” covalent inhibition .....	85
In silico approach for the design of selective covalent HCV inhibitors .....	86
Conclusion .....	91
CHAPTER 5 .....	99
Targeting HCV Polymerase: A Structural and Dynamic Perspective into the Mechanism of Selective Covalent Inhibition .....	99
Introduction.....	101
Computational Methodology .....	105
Results and Discussion .....	109
Conclusion .....	120
CHAPTER 6 .....	126
Enhanced Pharmacophore-based Virtual Screening Approach in the Discovery of Potential Inhibitors of HCV NS3 Helicase.....	126
Introduction.....	128
Computational Methodology .....	130
Results.....	133
Discussion.....	139
Conclusion .....	140

CHAPTER 7 .....	149
7.1 Conclusion .....	149
7.2 Future Perspective .....	150
APPENDICES .....	151

# CHAPTER 1

## 1 Introduction

### 1.1 Background and Rationale of the Study

Hepatitis C virus (HCV) is the main cause of hepatic-related diseases throughout the world. Approximately 3% of the world's population is believed to be chronic carriers of the virus (Karoney and Sika, 2013). Annually, at least 3 to 4 million individuals contract HCV, mainly through contaminated blood transfusions or intravenous drug use involving an infected person (Mohamed *et al.*, 2015; Thrift, El-Serag and Kanwal, 2017). The HCV infection is categorized into two states: acute and chronic. Usually the acute HCV infection is asymptomatic; however, in some cases, patients experience jaundice, malaise and slight fatigue (Blackard, 2008; Sagnelli *et al.*, 2014). Amongst those infected, 60-80%, which is approximately 130 to 170 million individuals will progress into the chronic state (Mohamed *et al.*, 2015). This state of infection is distinguished by the development of hepatic cirrhosis, fibrosis and steatohepatitis (El-Zayadi, 2008; Ghany and Jake Liang, 2016; Zeremski *et al.*, 2016). In cases of fatality, patients often succumb to the viral infection due to extensive hepatic damage caused by hepatocellular carcinoma development or the host's immune response (Baumert *et al.*, 2017).

There are two principle drug targets of HCV, non-structural protein 3 and 5B RNA-dependent RNA polymerase (NS3 and NS5B RdRp, respectively). The NS3 protein is a complex consisting of a serine protease, RNA helicase and co-factor, NS4A (Chevaliez and Pawlotsky, 2006; Shunmugam, Ramharack and Soliman, 2017). The protease portion of the complex is highly targeted for antiviral drug development owing to its vital role in the HCV life cycle. The protease functions by mediating the cleavage of precursor proteins for the generation of functional products and facilitating post-translational processing of structural proteins, required for viral assembly and morphogenesis. The NS3 RNA helicase is an often-overlooked aspect of the HCV enzyme complex. The helicase is considered to be an indispensable component of viral transcription, translation and RNA splicing; however, there is an evident lack in literature pertaining to the helicase and its possible role in antiviral drug discovery. The second most popular drug target is the NS5B RdRp. The protein is responsible for facilitating RNA genome transcription and subsequent viral replication (Eltahla *et al.*, 2015).

Presently, few antiviral drugs have been approved for HCV treatment, which targets the NS3 protease, such as narlaprevir (Arasappan *et al.*, 2010), telaprevir (Kwong *et al.*, 2011) and combined therapy with pegylated interferon-alpha and ribavirin ; and the RdRp, such as sofosbuvir (Bhatia *et al.*, 2014) and

dasabuvir (Goldenberg, 2014). However, there is still a pressing need for more effective HCV therapeutics. Numerous complications have been encountered, casting a negative stigma on the effectiveness of the antiviral drugs. It has been reported that only 50-60% of infected individuals positively respond to current HCV therapies (Bowen and Walker, 2005). Mutations due to the initiative of HCV to survive and escape the host immune response has facilitated the emergence of several genetic variations, thus giving way to drug resistance. Another serious drawback to current HCV drugs comes in the form of undesirable adverse effects, which harshly impacts the well-being of patients with an already compromised health status caused by the infection. To date, no preventive HCV vaccine is available, thus continual HCV research is an absolute necessity for the development of improved and effective antiviral therapies or vaccine against this ruinous disease. The major obstacles in achieving this goal is the absence of an appropriate small animal model susceptible to the HCV infection and the inability to cultivate HCV *in vitro*. The only animal that is capable of being infected by HCV is the chimpanzee, but due to ethical reasons, endangered species status and high maintenance expenses, its use is limited (Saleh, Elhaes and Ibrahim, 2017). These setbacks have prompted a surge in computational drug design research efforts for the development of antiviral therapies that selectively target HCV viral proteins and explore promising novel drug targets.

Computer-aided drug design (CADD) approaches have extensively contributed toward the design and development of drugs against a wide variety of diseases (Yu and MacKerell, 2017). Computational procedures are often utilised to accelerate the discovery process and direct attention toward compounds exhibiting potential antiviral activities, elaborate on protein-drug interactions and provide enhancements for eligible compounds (Kore *et al.*, 2012; Das, Saha and Abdul, 2017). The CADD strategy in drug discovery can be applied at every stage, from target identification, to optimisation of drugs and preclinical investigations (Lu *et al.* 2012; Huang *et al.*2010; Song *et al.* 2009; Anderson 2003). With diseases like HCV having a global affect, CADD techniques including molecular modelling, molecular docking, virtual screening, pharmacophoric moiety identification and molecular dynamic simulations allow medicinal and pharmaceutical industries to screen massive quantities of compounds to filtrate possible lead drugs that may then be experimentally validated. Implementation of CADD vanquishes the notion to estimate or guess with little to no assurance on the accuracy or reliability of experimental screening outcomes thus, reducing the time frame of drug discovery.

In this study, we have utilized major computational techniques to fill the gap in HCV drug design research, thus giving insights toward specific drug targets and designing potential selective inhibitors against Hepatitis C.

## **1.2 Aims and Objectives**

The aim of this thesis is to better describe and investigate the main target proteins of HCV and subsequently use CADD techniques to investigate potential small-molecule inhibitors against these antiviral proteins.

To achieve this, the following objectives were outlined:

1. To introduce a technical guidance for the design and development of potential selective covalent inhibitors against HCV NS3/4A by:
  - 1.1 Providing a comprehensive review on the target protein and covalent inhibition.
  - 1.2 Identifying and describing the importance of HCV NS3/4A cysteine159 in selective covalent inhibition.
  - 1.3 Creating a technical guideline that provides a systematic approach for the design and development of selective covalent inhibitors.
2. To better understand the structural and dynamic impact of selective covalent inhibitor binding to HCV NS5B RNA-dependent RNA polymerase (RdRp). This may be accomplished by:
  - 2.1 Utilising the crystal structure of HCV RdRp in complex with selective covalent inhibitor, compound 47, in a 200ns continuous molecular dynamic simulation.
  - 2.2 Performing post molecular dynamic analyses to investigate the underlying structural and dynamic behaviours of the apo RdRp in comparison to compound 47-bound RdRp.

- 2.3 Using additional defined parameters to detail the impact of ligand binding on RdRp, in its entirety and with respect to key components required for RNA binding and subsequent viral replication.
3. To computationally utilize the “Per-residue Energy Decomposition Pharmacophore” virtual screening approach to propose potential HCV NS3 helicase inhibitors using the chemical scaffold of quercetin as a template. This may be achieved by:
- 3.1 Performing molecular dynamic simulations to create an ensemble of potent inhibitors in complex with HCV NS3 enzyme.
  - 3.2 Quantification of each amino acid residue interaction toward the total binding free energy of the system, based on the MM/GBSA method. Thus, creating a pharmacophore-model of the complex erected from enzyme-ligand interactions.
  - 3.3 Initiating a search through the Zinc Database for structure-based scaffolds similar to that of quercetin, against the NS3 enzyme.
  - 3.4 Measuring the stability of the most favourable ligands as well as binding free energy of each complex following molecular dynamic simulation.
  - 3.5 To provide a preliminary investigation that serves as a fundamental platform for future wide-spectrum screening of potential HCV antiviral drugs against NS3 RNA helicase.

### **1.3 Novelty and Significance of the Study**

In the past decade, the drug design industry has made significant efforts toward developing a novel strategy that allows selective targeting of viruses, a concept known as selective covalent inhibition. This strategy employs structure-based approaches to generate small-molecule inhibitors that selectively bind to a specific residue of a target protein. The bond formed is designated as covalent and occurs between the ligand’s highly reactive warhead and the thiol group of a specific cysteine residue of a target protein. The cysteine residue must be conserved across all viral geno and subtypes, and not present in host enzymes, thus serving as a perfect candidate for acquiring selectivity. This strategy has been demonstrated to be promising to overcome efficacy, potency and selectivity challenges.

To date, there are no reported computational studies that provide information concerning this form of inhibition related to HCV and its viral proteins. The lack of literature regarding selective covalent inhibition, its underlying mechanism of action and the overall impact on structural dynamics of HCV prompted us to design a technical guideline or “road map” toward the design and discovery of potential selective covalent inhibitors for NS3/4A protease. The guideline was thereafter implemented using the NS5B RdRp to provide further insight into selective covalent binding mechanisms of potential antiviral drugs in HCV therapy. The study also reported a previously neglected viral drug target, the RNA helicase of the HCV NS3 protein. This was the first account of pharmacophore-based virtual screening using quercetin’s chemical scaffolding as a template against the NS3 helicase, thus assisting researchers from multidisciplinary fields in designing potential small-molecule inhibitors against the viral protein.

With the use of CADD techniques, a comprehensive *in silico* perspective is presented to elucidate possible structural characteristics that allow for the inhibition of these enzymes as well as amino acid residues implicated in activity of the enzyme. Highlighting the binding landscape will aid prospective design of selective, potent inhibitors with indispensable pharmacophoric features that will facilitate development of targeted and effective small-molecule inhibitors.

To this end, the work presented in this thesis is considered to be a fundamental platform in drug development research for the design of more potent and highly selective inhibitors against HCV NS3 and NS5B proteins with the potential to overcome challenges arising in current HCV therapeutics.

## Reference

Arasappan, A. *et al.* (2010) 'Discovery of narlaprevir (SCH 900518): A potent, second generation HCV NS3 serine protease inhibitor', *ACS Medicinal Chemistry Letters*, 1(2), pp. 64–69. doi: 10.1021/ml9000276.

Baumert, T. F. *et al.* (2017) 'Hepatitis C-related hepatocellular carcinoma in the era of new generation antivirals', *BMC Medicine*. doi: 10.1186/s12916-017-0815-7.

Bhatia, H. K. *et al.* (2014) 'Sofosbuvir: A novel treatment option for chronic hepatitis C infection.', *Journal of pharmacology & pharmacotherapeutics*, 5(4), pp. 278–284. doi: 10.4103/0976-500X.142464.

Blackard (2008) 'Acute hepatitis C virus infection: A chronic problem (Hepatology (2008) (321))', *Hepatology*. doi: 10.1002/hep.22205.

Bowen, D. G. and Walker, C. M. (2005) 'Adaptive immune responses in acute and chronic hepatitis C virus infection', *Nature*. doi: 10.1038/nature04079.

Chen, K. X. and Njoroge, F. G. (2010) 'The journey to the discovery of boceprevir: An NS3-NS4 HCV protease inhibitor for the treatment of chronic hepatitis C', *Progress in Medicinal Chemistry*, pp. 1–36. doi: 10.1016/S0079-6468(10)49001-3.

Chevaliez, S. and Pawlotsky, J.-M. (2006) 'HCV Genome and Life Cycle', in *Hepatitis C Viruses: Genomes and Molecular Biology*. doi: PMID: 21250393.

Das, P. S., Saha, P. and Abdul, A. P. J. (2017) 'A Review on Computer Aided Drug Design In Drug Discovery', *World Journal of Pharmacy and Pharmaceutical Sciences*, 6(7), pp. 279–291. doi: 10.20959/wjpps20177-9450.

El-Zayadi, A. R. (2008) 'Hepatic steatosis: A benign disease or a silent killer', *World Journal of Gastroenterology*. doi: 10.3748/wjg.14.4120.

Eltahla, A. A. *et al.* (2015) 'Inhibitors of the hepatitis C virus polymerase; mode of action and resistance', *Viruses*, 7(10), pp. 5206–5224. doi: 10.3390/v7102868.

Ghany, M. G. and Jake Liang, T. (2016) 'Natural history of chronic hepatitis C', in *Hepatitis C Virus II: Infection and Disease*. doi: 10.1007/978-4-431-56101-9\_1.

Goldenberg, M. M. (2014) 'Pharmaceutical approval update.', *P & T: a peer-reviewed journal for formulary management*, 37(9), pp. 499–502. doi: 10.1136/ebmh.10.1.16.

Karoney, M. J. and Sika, A. M. (2013) 'Hepatitis C virus (HCV) infection in Africa: a review', *Pan African Medical Journal. Pan African Medical Journal*, 141444, pp. 44–1937. doi:



10.11604/pamj.2013.14.44.2199.

Kore, P. P. *et al.* (2012) ‘Computer-Aided Drug Design: An Innovative Tool for Modeling’, *Journal of Medicinal Chemistry*, 2(1), pp. 139–148.

Kwong, A. D. *et al.* (2011) ‘Discovery and development of telaprevir: an NS3-4A protease inhibitor for treating genotype 1 chronic hepatitis C virus.’, *Nature biotechnology*, 29(11), pp. 993–1003. doi: 10.1038/nbt.2020.

Mohamed, A. A. *et al.* (2015) ‘Hepatitis C virus: A global view’, *World Journal of Hepatology*. doi: 10.4254/wjh.v7.i26.2676.

Sagnelli, E. *et al.* (2014) ‘Hepatic flares in chronic hepatitis C: Spontaneous exacerbation vs hepatotropic viruses superinfection’, *World Journal of Gastroenterology*. doi: 10.3748/wjg.v20.i22.6707.

Saleh, N. A., Elhaes, H. and Ibrahim, M. (2017) *Design and development of some viral protease inhibitors by QSAR and molecular modeling studies, Viral Proteases and Their Inhibitors*. Elsevier Inc. doi: 10.1016/B978-0-12-809712-0.00002-2.

Shunmugam, L., Ramharack, P. and Soliman, M. E. S. (2017) ‘Road Map for the Structure-Based Design of Selective Covalent HCV NS3/4A Protease Inhibitors’, *Protein Journal*. Springer US, 36(5), pp. 397–406. doi: 10.1007/s10930-017-9736-8.

Thrift, A. P., El-Serag, H. B. and Kanwal, F. (2017) ‘Global epidemiology and burden of HCV infection and HCV-related disease’, *Nature Reviews Gastroenterology and Hepatology*. doi: 10.1038/nrgastro.2016.176.

Yu, W. and MacKerell, A. D. J. (2017) ‘Computer-Aided Drug Design Methods.’, *Methods in molecular biology (Clifton, N.J.)*. NIH Public Access, 1520, pp. 85–106. doi: 10.1007/978-1-4939-6634-9\_5.

Zeremski, M. *et al.* (2016) ‘Fibrosis Progression in Patients with Chronic Hepatitis C Virus Infection’, *Journal of Infectious Diseases*. doi: 10.1093/infdis/jiw332.

## CHAPTER 2

### 2. Background on the Hepatitis C Virus

#### 2.1 Introduction

Hepatitis C Virus (HCV) is a major global health concern and is the main cause of hepatic-related diseases such as hepatitis C, cirrhosis, fibrosis and hepatocellular carcinoma (Ghany and Jake Liang, 2016). Classified as hepatotropic RNA virus, HCV belongs to *Hepacivirus* genus and is a member of the *Flaviviridae* family (Lavanchy, 2011; Villani, Bellanti and Serviddio, 2013). The World Health Organisation (WHO) revealed that HCV is responsible for affecting approximately 3% of the world's population (Lavanchy, 2011; Karoney, Siika and Karoney, 2013; World Health Organization, 2018). Every year, 3 to 4 million new HCV infections occur while 170 to 185 million of infected individuals will progress to the chronic state (Mohamed *et al.*, 2015; Thrift, El-Serag and Kanwal, 2017). Currently, there is no HCV vaccine available and to make matters worse, compromised health of infected individuals and antiviral resistance facilitated by increased presence of viral genetic variants have greatly dampened the success rate of market-available drugs (Fonseca-Coronado *et al.*, 2012; Echeverria *et al.*, 2015; Kumthip and Maneekarn, 2015). For this reason, constant HCV research is essential for the development of selective antiviral therapies or vaccines against this devastating disease.

In this chapter, we provide a comprehensive review of HCV that describes its discovery, epidemiology, transmission, life cycle and overall molecular virology. The main focus will be on HCV non-structural protein 3 (NS3) and 5B (NS5B) as therapeutic drug targets for the design of selective antiviral compounds, which is the key theme of the thesis research.

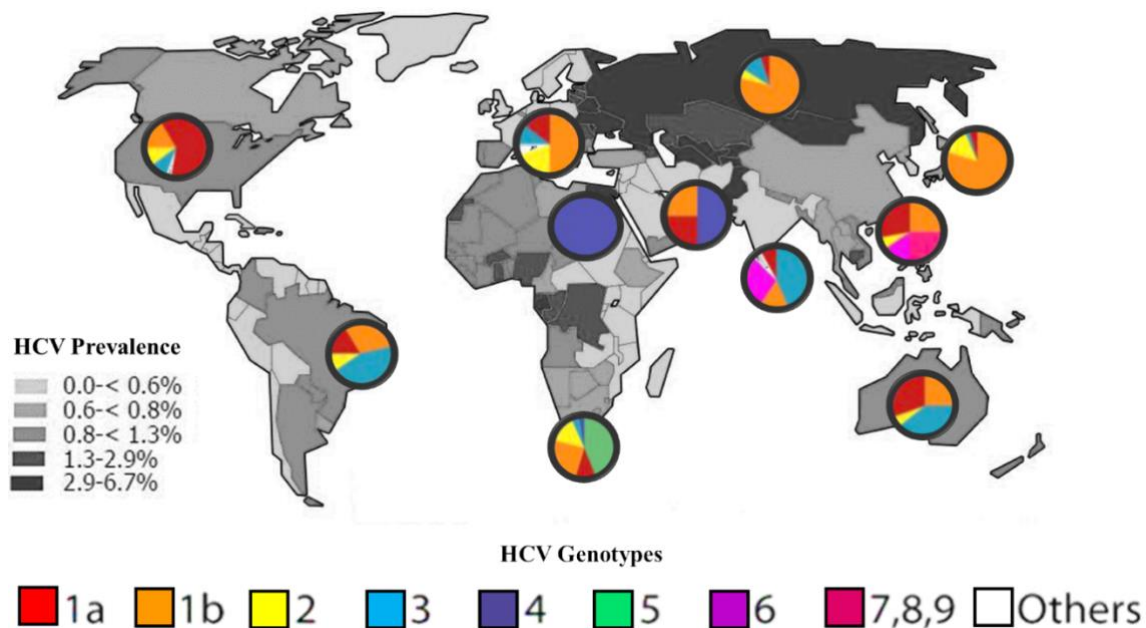
#### 2.2 Discovery and Epidemiology

The rigorous task of diagnostic assay development for hepatitis A and B virus (HAV and HBV, respectively) led Feinstone and co-investigators to discover an additional viral hepatitis of unknown origin (Feinstone *et al.*, 1975, 1981). In later years, the unknown viral hepatitis was given the name non-A, non-B (NANB) hepatitis as majority of cases involving transfusion-associated hepatitis were not associated with HAV or HBV infections (Feinstone *et al.*, 1975; Bukh, 2016). With the timely development of powerful molecular tools and techniques, the etiological agent of transfusion-associated hepatitis was identified. The breakthrough in

identifying the virus occurred in the late 1980s whereby Choo and co-researchers isolated antibodies from an individual infected with the unknown viral hepatitis, for the detection of specific viral proteins (Choo *et al.*, 1991). Nucleic acid was extracted from the plasma of Chimpanzee and experimentally transfected with NANB hepatitis. Thereafter, the infected sample was subjected to reverse transcription for complementary DNA (cDNA) generation. Using the cDNA, restriction enzyme fragments were inserted into a bacteriophage expression vector. The resultant proteins expressed were then analysed by comparing them with the expressed viral proteins from the serum of the patient infected with the unknown viral agent (Choo *et al.*, 1991). As a result, the causative agent of transfusion-associated NANB hepatitis was successfully identified and designated the named, hepatitis C virus (HCV).

Globally, HCV is one of the leading causes of mortality and morbidity (Fedeli *et al.*, 2017). In the last ten years, there has been a 2.8% increase in the number of individuals who have tested positive for the infectious agent (Petruzzello *et al.*, 2016). Currently, it is estimated that approximately 185 million individuals are infected with HCV worldwide (Thrift, El-Serag and Kanwal, 2017). The acute stage of HCV infection is relatively slow, asymptomatic and spontaneously clears up within six months without medical intervention. Unfortunately, an estimated 60-80% of HCV-infected individuals will progress to the chronic state of infection (World Health Organization, 2018). The chronic state of HCV infection is commonly linked to the development of hepatocellular cancers (Goossens and Hoshida, 2015), cirrhosis (Khullar and Firpi, 2015), fibrosis (Sebastiani, Gkouvatsos and Pantopoulos, 2014), hepatic failure and ultimately death (Toshikuni, Arisawa and Tsutsumi, 2014). To date, there is no vaccine available for use, therefore infected patients heavily rely on therapeutic interventions to treat the disease and alleviate symptoms but its only effective in 50% of individuals under treatment.

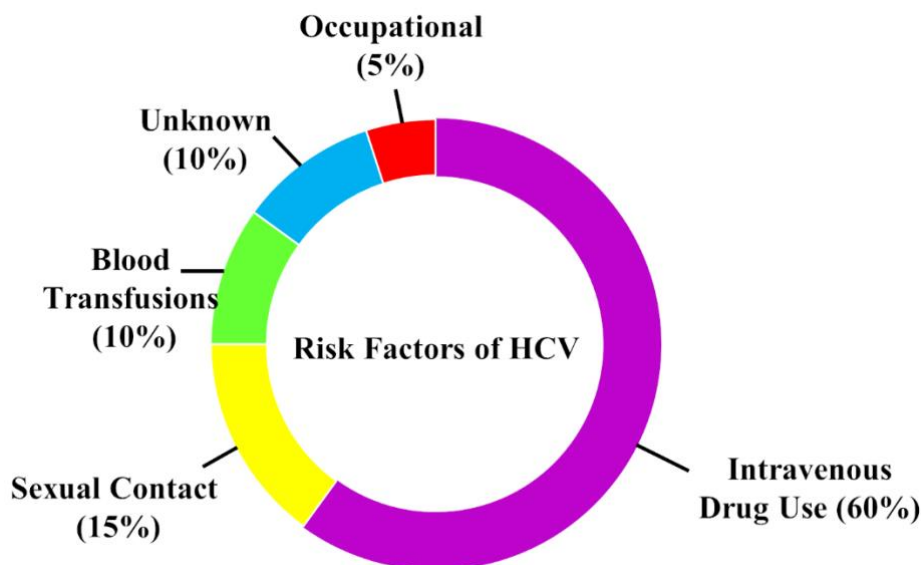
A high degree of genetic diversity observed in HCV has resulted in the emergence of nine distinct genotypes and numerous subtypes of the virus. Despite sharing similar fundamental virology, the genotypes differ based on their geographical location and rate of incidence. The geographical distribution of HCV is represented in Figure 1. A unique feature of HCV is that the virus exists as a populace of non-identical but closely related “quasispecies”, which emerge from mutations accruing over time and are not considered different subtypes. The generation of multiple variants are due to the error-prone nature of HCV as it lacks adequate proof-reading activities during the replication process.



**Figure 1:** Geographical representation of global HCV genotype distribution (Adapted from Gower *et al.*, 2014).

### 2.3 Route of Transmission

The major risks of infection with HCV is through direct percutaneous exposures including sharing needles with infected individuals, medical surgery such as organ transplantation from an infected donor and receiving infected blood during transfusions (Alter, 2007; Thursz and Fontanet, 2014). Individuals at high risk of contracting the virus through blood are those that require continuous blood transfusions due to their infliction with diseases such as cancer, haemodialysis-associated with chronic renal failure, haemophilia and thalassemia (Moosavy *et al.*, 2017). Other less efficient transmission routes include mucosal exposure to serum or blood during health-care associated accidental needlestick injuries (Pozzetto *et al.*, 2014), sexual intercourse (Wuytack *et al.*, 2018) or childbirth (Tosone *et al.*, 2014). Perinatal transmission of HCV occurs only in instances involving maternal patients with high viral loads or membrane rupturing during prolonged labour (Alter, 2007). Another reservoir of HCV transmission is unintentional percutaneous exposures that can cause infection through cross-contamination. This involves contact with injection-type drug paraphernalia as well as inefficient sterilisation or reuse of medical equipment, for example, infusion bags, vials, syringes and needles (Alter, 2007; Kaushik, Kapila and Praharaj, 2011). The contributing risk factors and their associated incidence rate in percentage are represented in Figure 2.

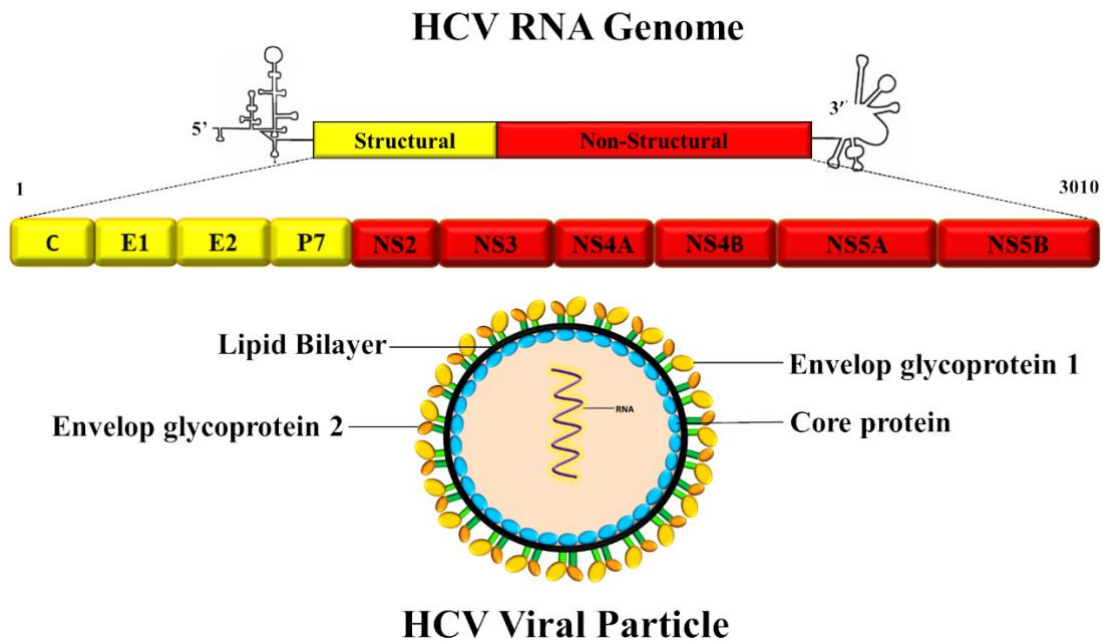


**Figure 2:** Risk factors associated with transmission of HCV (Adapted from Stambouli, 2014).

## 2.4 Molecular Virology

### 2.4.1 Genome Organisation

The HCV genome is represented by a positive sense, single-stranded RNA, approximately 9600 nucleotides in length. The genome is made up of a single open reading frame (ORF), flanked by the 5' and 3' untranslated regions (5'UTR and 3'UTR, respectively), which is vital for genome replication and polyprotein translation (Figure 3) (Kalinina and Dmitriev, 2015). The ORF encodes the polyprotein, which undergoes proteolytic cleavage by cellular and viral proteases to generate 10 viral proteins (Takamizawa *et al.*, 1991; Chevaliez and Pawlotsky, 2006; Ashfaq *et al.*, 2011; Kalinina and Dmitriev, 2015). The structural proteins are the core (C), envelope protein 1 and 2 (E1/E2) and hydrophilic ion channel protein 7 (P7), they are located at the N-terminal of the polyprotein. The remaining two thirds of the HCV genome is designated as non-structural (NS) proteins: (NS2, NS3, NS4A, NS5A and NS5B) (Villani, Bellanti and Serviddio, 2013; Kalinina and Dmitriev, 2015).



**Figure 3:** Schematic representation of hepatitis C virus (HCV) genome organisation of the polyprotein and viral particle (Prepared by author).

#### 2.4.2 Viral Proteins

##### 2.4.2.1 Structural

The first structural protein encoded by the HCV ORF is the core (C), which is responsible for nucleocapsid formation and viral morphogenesis (Gawlik and Gallay, 2014). This is a multifunctional protein incriminated in the pathogenesis of HCV. Localisation of the C protein occurs at the endoplasmic reticulum (ER) membrane, on lipid droplet surfaces and in the membranous network. The latter is a structure formed by host factors and viral proteins of HCV. The protein is also linked to several other processes such as cell proliferation and death, regulation of gene transcription, suppression of host immune response and lipid metabolism interference (Moradpour and Penin, 2013; Adler, 2014).

The HCV polyprotein encodes two envelope glycoproteins, E1 and E2. These proteins serve as building blocks for the viral envelope and is processed by host peptidases. Expression of both glycoproteins are crucial for viral entry and heterodimer formation on the viral envelop (Lavie, Goffard and Dubuisson, 2007). The E2 protein encompasses a hyper-variable domain that undergoes frequent mutations to sanction the development of antigenic escape variations. In doing so, the host's immune response can no longer recognise the virus, thus elimination is not

possible (Law *et al.*, 2018). Both, E1 and E2 have vital functional roles in viral morphogenesis, assembly and entry (Lavie, Goffard and Dubuisson, 2007; Vieyres, Dubuisson and Pietschmann, 2014).

The p7 protein is identified as hydrophobic peptide with cation-selective ion channel activity (Chandler *et al.*, 2012; Madan and Bartenschlager, 2015). The p7 protein is relatively small, consisting of 63 amino acids and is partially cleaved by host signal peptidase, from the E2 protein (Jones *et al.*, 2007; Madan and Bartenschlager, 2015). Even though the P7 protein is not essential for RNA replication, it is however important in viral particle production and subsequent release (Lohmann *et al.*, 1999; Jones *et al.*, 2007; Gentzsch *et al.*, 2013).

#### 2.4.2.2 Non-Structural

The NS2 transmembrane protein that contains an auto-protease that conducts cleavage at the NS2/3 juncture (Ma *et al.*, 2011). This cleavage is an essential step in viral replication of HCV by permitting downstream processing of the polyprotein. Additionally, the NS2 protein has also been known to have an important role in assembly of the virion and it interacts with the envelope proteins (Stapleford and Lindenbach, 2011).

The NS3 protein contains two enzymatic functions: helicase and serine protease. The helicase activity is actively involved in assembly of the HCV virion and the protease activity is responsible for HCV polyprotein cleavage, in order to generate vital proteins required for RNA genome replication (McGivern *et al.*, 2015). The co-factor, NS4A aids in membrane association of the NS3 protein (Kim, Han and Ahn, 2016). The NS3 protein is one of the major focuses of this work, the structural expression, characteristics, enzymatic activity, and interaction with other host and viral factors as well as the efficiency of anti-viral compounds is further elaborated in more detail in subsequent sections.

NS4B is a small polytopic membranous protein, whose role is not well understood but it is thought to serve as a frame for replicase assemblage (Paul *et al.*, 2011; Kazakov *et al.*, 2015). This protein has created quite a challenge in its investigation as it exhibits relatively high hydrophobicity (Palomares-Jerez, Nemesio and Villalaín, 2012). The protein is subdivided into three domains: A cytosolic C- and N-terminal as well as a central domain. At the N-terminal, the protein fosters two amphipathic helices (AH1 and AH2), AH2 is known to contribute to NS4B association to cellular membranes (Blight, 2011; Gouttenoire *et al.*, 2014; David *et al.*, 2015). The NS4B protein has three main functional features that are perceived to be contributing factors toward initiating

the formation of the membranous web. Firstly, NS4B has the ability to form oligomers, which is facilitated mainly by C-terminal components and amphipathic helix 2. The second feature is the ability of the protein's amphipathic helix 2 domain to endorse aggregation of lipid vesicles. Lastly, the third feature is the ability of the N-terminal to post-translationally translocate into the ER lumen, from the ER cytosolic domain. It was also found that this translocation is negatively impacted by NS5A, suggesting that NS5A influences NS4B topology (David *et al.*, 2015).

NS5A is a zinc-binding, membrane metalloprotein. This protein exhibits pleiotropic functions including viral replication, virion assembly, and extensive interactions with cellular processes such as tumorigenesis promotion and apoptosis inhibition, both of which trigger hepatocarcinogenic progression (Kim and Chang, 2013; Pawlotsky, 2013). The NS5A phosphorylation status is considered to have a pronounced impact on its enzymatic function and the HCV life cycle. The phosphorylation states of the protein are thought to mediate the efficiency of HCV replication, seemingly through interactions with other HCV NS proteins or regulatory factors of the host. This protein is found in two forms: hyperphosphorylated (p58; 58 kDa) and basally phosphorylated (p56; 56 kDa), which are each implicated in different functions (Pawlotsky, 2013; Masaki *et al.*, 2014; Masaki and Suzuki, 2015). The NS5A protein is subcategorised into three domains: domain I, II and III. The NS5A protein is anchored to the membrane through an amphipathic helix found at the N-terminal of domain I. To date, the functions of domain II and III are still unknown (Scheel and Rice, 2013).

The NS5B RNA-dependent RNA-polymerase (RdRp) of HCV is an essential component of the RNA replicase and is responsible for replication of the RNA genome (Han *et al.*, 2014). The HCV RdRp protein serves as another main focus area of this work. Its structural features, expression, enzymatic activity, interactions with other virus-associated and host factors and the effectiveness of anti-viral compounds is deliberated in more detail in subsequent sections.

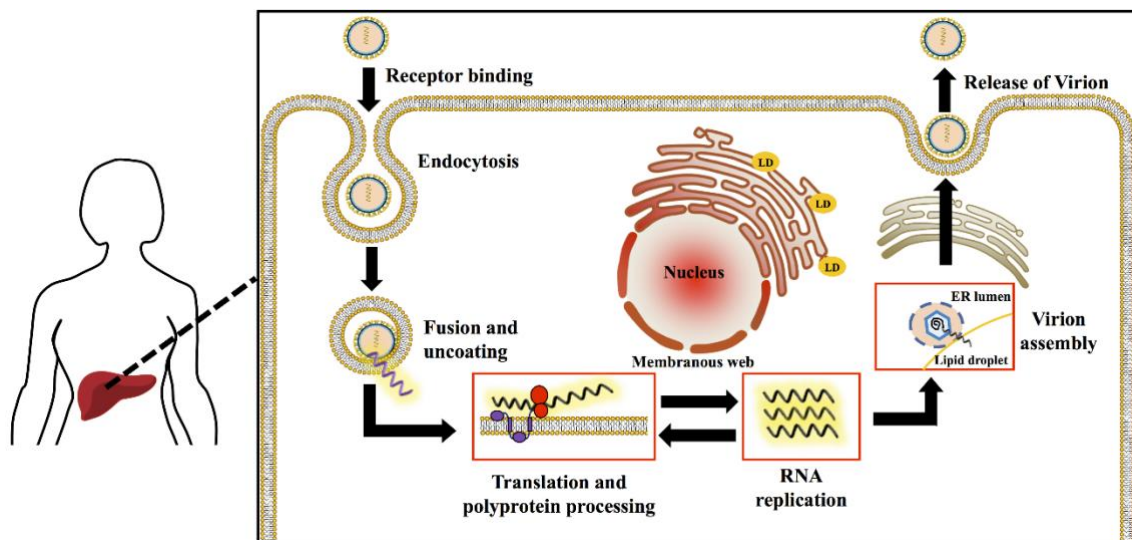
Among all the HCV proteins, NS3 and NS5B proteins serve as the most attractive and potentially aspiring targets in drug design and discovery research.

#### 2.4.3 Life cycle of HCV

The HCV is a blood-transmitted virus that reaches the liver via the circulatory system (Irshad, Mankotia and Irshad, 2013). The HCV virion binds to a viral receptor present at the surface of the host liver cell. This binding facilitates the internalisation of the viral particle through fusion of the viral envelop and the cellular membrane via clathrin-mediated endocytosis (Kim and



Chang, 2013). Once the virus gains entry to the cell, the viral particle undergoes uncoating to initiate the release of the HCV positive sense RNA. The RNA in question, encompasses an internal ribosome entry site (IRES) located in the 5'-end. This allows mimicking of the cellular mRNA 5'-cap, thus duping cellular ribosomes to induce translation into a polyprotein. The polyprotein is associated with the cellular endoplasmic reticulum (ER) and is processed by viral and cellular proteases into structural and non-structural HCV proteins. The HCV proteins form a replication machinery, which create RNA copies in a part of the ER called the membranous web. The newly synthesised RNA-molecules are subsequently encapsulated and gathered into new virion particles that mature prior to leaving the cell by means of exocytosis (Dustin *et al.*, 2016). The life cycle of HCV is diagrammatically presented in Figure 4.



**Figure 4:** Overview of HCV life cycle in liver cells (Adapted from Wang *et al.*, 2017).

## 2.5 Clinical Features

### 2.5.1 Acute HCV Infection

Upon infection with HCV, majority of patients will enter the acute phase. Generally, acute HCV infection spontaneously resolves within six months and is clinically mild or asymptomatic. Since this phase presents with no symptoms, its diagnosis and study with respect to the host immune response and viral clearance is an extremely difficult task. This is the reason as to why many HCV-infected patients are diagnosed only in the chronic state of infection and the time of exposure cannot be accurately pin-pointed (Butler *et al.*, 2016). Following exposure, HCV-RNA can be detected in the host's serum within 1 to 3 weeks. Symptoms that may occur include: loss of appetite, abdominal pain, dyspepsia, jaundice and fatigue (Deutsch *et al.*, 2013; Butler *et al.*,

2016). In most cases, the symptoms are not severe and do not require medical treatment. Elevated alanine aminotransferase (ALT) levels are the first sign of hepatic injury and can be detected 4 to 12 weeks after contracting the virus (Choi *et al.*, 2016). The HCV infection is detected by the presence of the viral RNA in the host's serum and through anti-HCV antibody seroconversion (Li and Lo, 2015). Clearance of the infection is partly influenced by host factors such as age, race, gender and human leukocyte antigen system; and viral factors such as genotype and co-infections present. The key driving force facilitating viral clearance of HCV infection is a strong cellular immune response of the host. It is important to note that patients who undergo spontaneous clearance should be monitored as relapse may occur up to 6 months after viral eradication.

### *2.5.2 Chronic HCV Infection*

The leading cause of hepatic-related diseases is HCV. As mentioned, a rather small percentage of individuals have the capability to eradicate the virus following an acute infection, regrettably, the majority will progress to the chronic state. Chronic HCV infection may persist for several years without the occurrence of death. Unfortunately, in most cases progression of the infection leads to hepatitis, fibrosis and cirrhosis development. At least 1-4% of chronically infected patients will develop HCC. Host-induced factors that intensify progression of hepatic damage include: hepatitis B virus (HBV) or human immunodeficiency virus (HIV) co-infection, race, gender, alcohol consumption, age and obesity (Yan and Wang, 2017). It appears that the HCV genotype does not influence progression of the disease, but non-alcoholic fatty liver disease can significantly worsen the state of the chronic viral infection (Silini *et al.*, 1996; Patel and Harrison, 2012).

## **2.6 Pathogenesis**

In the pathogenesis of viral hepatitis, the host immune response has a peculiar role as it modulates and attenuates the infection as well as contributing toward development of the chronic infection and hepatic cirrhosis (Park and Rehmann, 2014). Viruses are categorised as cytopathic, meaning cells are killed during the infection, or non-cytopathic such as HCV (Plesa *et al.*, 2006). The HCV induces acute or chronic hepato-related diseases whilst intermingling with the host's immune system.

In the first few weeks following exposure, interactions between the host immune system and HCV may significantly influence the viral infection in terms of its evolution and prognosis. The progression of HCV infection is usually slow and may take up to twenty years before the effects

on the host are seen. Generally, the acute stage of the infection is asymptomatic and on rare occasions, may be associated with life-threatening illnesses. The acute HCV infection accounts for 15-45% of the infected individuals and will clear up within six months without medical intervention, unfortunately, an estimated 60-80% will go onto develop chronic HCV infection (World Health Organization, 2018).

Following exposure to HCV, the innate immune response is triggered but the virus has the capability to comprehensively evade the response, which leads to viral persistence (Li and Lemon, 2013). This phenomenon takes place due to the evolution of HCV that allows the offset of retinoic acid-inducible gene I (RIG-1) pathway, thereby allowing the virus to elude the immune-induced challenge. This occurrence is the reason behind the majority of HCV-infected individuals progressing to the chronic state of infection. During this process, the protease domain of the NS3 protein is activated for cleavage of interferon-beta (IFN- $\beta$ ) promoter stimulator (IPS-1), an adapter protein of RIG-1 (Park *et al.*, 2016). Following cleavage, IPS-1 is not able to generate downstream signals for activation of nuclear factor kappa-light-chain-enhancer of activated B (NF $\kappa$ B) cells and interferon regulatory transcription factor 3 (IRF-3), resulting in the incapability of HCV-infected cells to express interferon-stimulated genes or produce IFN- $\beta$  (Irshad, Mankotia and Irshad, 2013).

Natural killer (NK) cells are key components, functional in the innate immune system. These cells have an important role in HCV eradication. A large population of NK cells can be found in the liver and are usually activated in the early phase of HCV infection. Activation of these cells leads to the recruitment of T lymphocytes and prompts hepatic antiviral immunity (Golden-Mason and Rosen, 2013; Holder, Russell and Grant, 2014). Additionally, the NK cells eliminate HCV-infected liver cells by secreting cytokines such as tumor necrosis factor-alpha (TNF- $\alpha$ ) and IFN- $\gamma$  or through cytolytic mechanisms. Unfortunately, HCV has adopted multiple defense strategies to counteract the response emanating from host NK cells. Compromised or inactive NK cells permit continuation of viral invasion whereas the active form of these cells contributes to hepatic injury (Irshad, Mankotia and Irshad, 2013). Through immunological assessments, it was established that onset of the adaptive cell-mediated and humoral immune response is delayed by 1-2 months and 2-3 months, respectively. This imminent delay allows HCV to bypass the adaptive immune system (Căruntu and Benea, 2006; Dustin *et al.*, 2016).

Liver damage caused by HCV exist primarily in the form of inflammation. This condition is known as hepatitis. Chronic persistence of hepatitis leads to fibrosis (Sebastiani, Gkouvatsos and Pantopoulos, 2014). When hepatic inflammation occurs, stellate cells are activated. Under normal

physiological conditions, these cells regulate blood flow through the liver and stores vitamin A (vit A) and fat (Wang *et al.*, 2013). Inflammation induced by HCV ignites reactions from all different types of hepatocytes, this causes stellate cells to alter their function from storing vit A to releasing the nutrient. During this process, cytokines are released, which in turn activates white blood cells (WBC) located externally to the liver. The WBCs migrate to the site of infection where they associate with Kupffer cells to yield chemical signals (Fujita and Narumiya, 2016). Kupffer cells are specialised WBCs that function to neutralise and eradicate tumorigenic cells and pathogens from the liver. These signals trigger collagen production by stellate cells within the extracellular matrix, an area suited between cells (Dixon *et al.*, 2013). Collagen is a fibrous protein, the main component in scar tissue formation (Bataller and Brenner, 2005). Normally, the body utilises collagen at the site of infection to restrict dissemination of the infection to neighbouring cells. As the infection resolves, the collagen matrix encapsulating the injury is dissolved and activated stellate cells undergo apoptosis to allow the tissue to return to normality. In circumstances where chronic HCV infection is present, the collagen matrix grows at a rapid, uncontrollable rate and cannot be efficiently dissolved. This results in accumulation of scar tissue around the cells. Eventually, hard, irregular areas of scar tissue, referred to as nodules will form and replace the smooth liver. The nodules disrupt blood flow to the hepatocytes causing the loss of valuable nutrients and oxygen that they require to survive, ultimately leading to cell death. A positive feedback mechanism between inflammation and fibrogenesis will cause amplification of fibrosis (Greener, 2013).

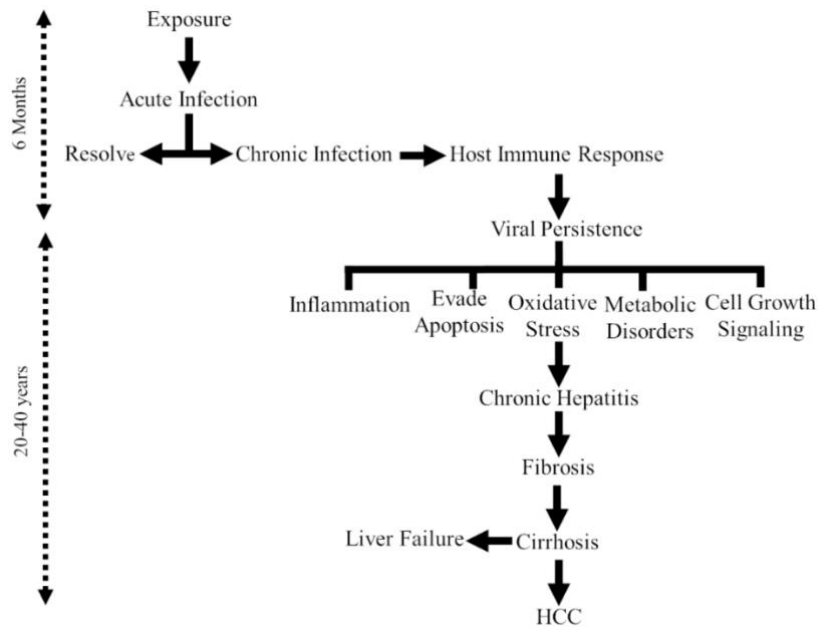
Oxidative stress may also be a possible cause of fibrosis due to the generation of free radicals (FRs), which are the by-product of cellular metabolic processes such as lipolysis and energy generation (Ivanov *et al.*, 2013; Phaniendra, Jestadi and Periyasamy, 2015). During these cellular processes, oxygen is converted into FR superoxide (Phaniendra, Jestadi and Periyasamy, 2015). Healthy, normal cells possess protective mechanisms that act against FRs; however, increased production of FR will inflict damage to tissues, cellular macromolecule including proteins, nucleic acids and lipids (Rahal *et al.*, 2014). In chronic HCV infection, prolonged inflammation may lead to FR over-production within the liver. The FRs will then attack the membranes of cells causing injury and cell death, thus further facilitating the progression of fibrosis (Ivanov *et al.*, 2013; Medvedev, Ploen and Hildt, 2016).

Experimental studies have positively established a link between type 2 diabetes mellitus (T2DM) and HCV, patients with chronic HCV infection frequently present with this form of diabetes (Hum

and Jou, 2018). Insulin resistance (IR) is a condition in which cells are unable to effectively respond to the normal action of insulin, thus becoming resistant to the hormone (Bose and Ray, 2014). In T2DM, IR is a common occurrence, coupled together, both significantly influence the state of the chronic HCV infection by enhancing development and progression of hepatic fibrosis and steatosis (Kralji *et al.*, 2016). Many chronically infected HCV patients are noted to have steatosis. Hepatic steatosis is described as the excessive accumulation of lipids within the liver. Two forms of steatosis are present in HCV-infected individuals: (1) metabolic steatosis and (2) HCV-induced steatosis (Adinolfi, Restivo and Marrone, 2013). Metabolic steatosis occurs due to IR, hyperlipidaemia and obesity (Kralji *et al.*, 2016). This form of steatosis is not triggered by HCV but its concurrence with the virus has been associated with rapid fibrosis progression. Metabolic steatosis can occur in all genotypes of HCV. The second form of steatosis is caused only by HCV genotype 3; however, the precise mechanism remains elusive, but it is understood that the virus directly induces steatosis through three mechanisms: impaired lipid secretion, elevated de novo free fatty acid synthesis and impaired degradation of fatty acids (Adinolfi, Restivo and Marrone, 2013; Irshad, Mankotia and Irshad, 2013). Cumulatively, this will result in the excess abundance of lipids that will further contribute to the steatosis endured by the patient. Upon treatment, HCV-induced steatosis regresses and may be entirely cleared but if relapse takes place, HCV-induced steatosis will reappear (Irshad, Mankotia and Irshad, 2013).

The development of HCV-induced HCC can span over a 20 to 40-year period and will only develop in approximately 1-4% of infected individuals with chronic infection. Carcinogenesis inflicted by HCV is mediated by the host-induced immunological response and viral-induced factors (Goossens and Hoshida, 2015; Axley *et al.*, 2018). Viral proteins of HCV are capable of directly interacting with cell signaling pathways to promote HCC development by upregulating cell division and growth or by inhibiting cell-cycle checkpoints and tumour suppressor genes (Hoshida *et al.*, 2014). In addition to HCV, the host-immunological response is also modulated by chronic inflammation, IFNs and TNF (Yu *et al.*, 2012). Accelerated rates of cell cycle repetition is often linked to the accumulation of mutations that may ultimately facilitate hepatocyte transformation into cancerous cells. The most common genes that undergo mutation in HCC are  $\beta$ -catenin, tumour protein 53 (p53) and telomerase reverse transcriptase, these mutations compromise the efficiency of telomerase maintenance thus leading to the upregulation of oxidative stress leading to cell death and regeneration, which gives rise to hepatocyte mutations and further supports HCC development (Nault *et al.*, 2014; Axley *et al.*, 2018). As cells begin to transform into cancerous cells, they undergo increased proliferation that causes the telomeres to

become very short. In normal cells, if telomeres become too short, they signal arrest of proliferation, cells become inactive and die (Carulli and Anzivino, 2014). Contrastingly, malignant cells abscond apoptosis by producing more telomerase enzymes, which stops the telomeres from getting even shorter (Donate and Blasco, 2011). The emergence of HCV-induced HCC is caused by the aggregation of cirrhotic hepatocytes that harbour excessive mutations, which allow readmittance into the cell cycle and reactivation of telomerase, thus permitting their advancements through cancer checkpoints, promoting progression of HCC development. The development of cirrhosis and fibrosis greatly increases the risk of developing HCC (Axley *et al.*, 2018). The pathogenesis of chronic HCV infection is summarised in Figure 5.



**Figure 5:**  
of HCV  
pathogenesis (Prepared by Author).

Overview

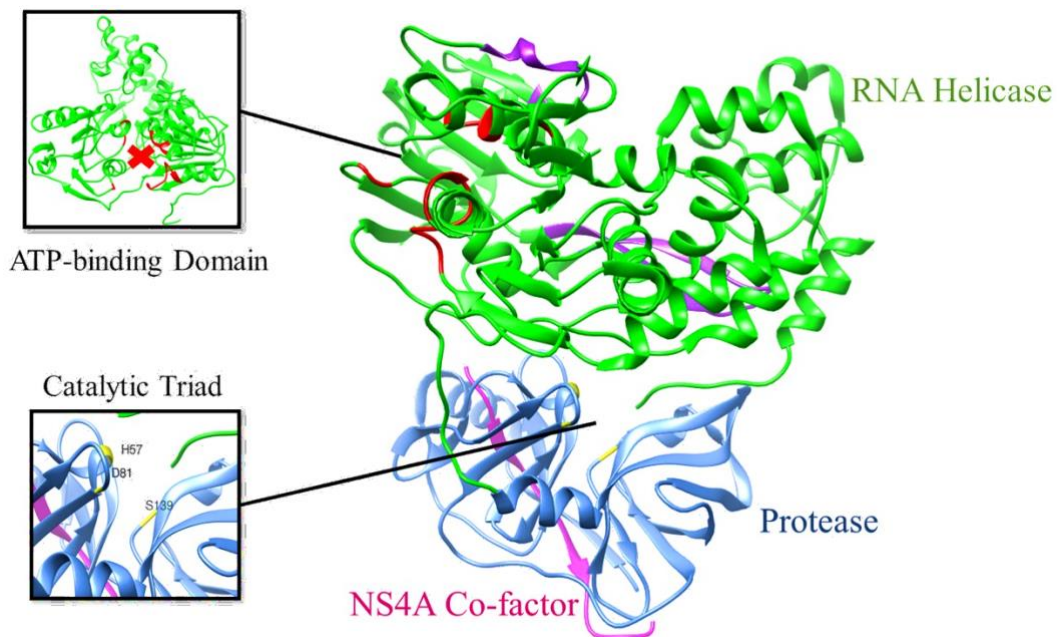
## 2.7 HCV NS3

### 2.7.1 Structure, Function and Potential as a Drug Target

The NS3 protein is a multifunctional viral enzyme, one third of the protein consists of a serine protease located at the N-terminal and the remaining two-thirds is an RNA helicase, positioned at the C-terminal of the genome. The NS3 chymotrypsin-like serine protease is accountable for downstream protein cleavage of junctions located between NS3/4A, 4A/4B, 4B/5A, 5A and 5B (Salam and Akimitsu, 2013). The protease contains the catalytic triad made up of three residues:

histidine 57, aspartate 81 and serine 139 that is required for efficient protease activity. The proper orientation the catalytic triad is dictated by the NS4A protein. This protein binds to the protease as a co-factor to stabilise and activate protease activity as well as to promote anchorage of NS3 to host cell membranes (Shiffman, 2012; Rimmert *et al.*, 2014).

The RNA helicase is the larger portion of the NS3 complex and is classified into the superfamily 2, DEXH-family of helicases. The function of the helicase is to separate the double-stranded RNA, using adenosine triphosphate as a molecular energy source (Byrd, 2012). Additionally, the helicase contains two important structural elements, the arginine clamp and the phe-loop. These two sequence motifs are vital for RNA or DNA binding and strand separation. Remarkably, these two regions are conserved across all known geno- and sub-types of HCV and they cannot be found in other SF2 protein families (Lam, Keeney and Frick, 2003; Frick, 2006). The NS3/4A complex and essential structural features are highlighted in Figure 6.



**Figure 6:** The three-dimensional crystal structure of HCV NS3 protein complex (PDB ID: 4B6E) highlighting important elements: the RNA helicase (green), protease (blue), NS4A (magenta), Arg-clamp and Phe-loop (purple), catalytic triad and the ATP-binding domain using UCSF Chimera software program. (Prepared by Author).

The NS3 complex is essential for HCV polyprotein cleavage, in order to generate vital proteins required for RNA genome replication (McGivern *et al.*, 2015). Within infected cells, the NS3 protein is also known to impede signaling of the innate immune recognition and response through cleavage of TIR-domain-containing adapter-inducing interferon- $\beta$  (TRIF) and mitochondrial

antiviral signaling (MAVS) proteins. The TRIF and MAVS proteins act in response to downstream stimuli received from the toll-like-receptor 3 and retinoic acid inducible gene-I pathways, respectively, leading to curtailment of the host cellular interferon response (Brown *et al.*, 2016; Liang *et al.*, 2018). The NS3/4A enzyme exhibits several functions including: unravelling of the HCV RNA genome from the RNA binding proteins, unwinding the single-stranded RNA with secondary structures and double-stranded (ds) RNA replication intermediates (Shiryaev *et al.*, 2012; Moradpour and Penin, 2013). Therefore, the NS3/4A enzyme is recognized as an indispensable component of the HCV replication complex (Salam and Akimitsu, 2013). The NS3 protein complex modulates both the host and virus, therefore, it is considered as an attractive target for drug development (McGivern *et al.*, 2015).

Over the years, the scientific community has focussed mainly on the protease as a drug target. This strategy has been effective so far, but it has also cast a relatively large shadow over the potential of the HCV helicase in drug therapy. The helicase is highly essential for transcription, translation and RNA splicing during replication. Not only is it pivotal in replication but it also contains two highly conserved regions, the Arg-clamp and Phe-loop that are exclusive only to HCV helicase (Lam, Keeney and Frick, 2003). Thus, the RNA helicase of HCV reveals itself as an addition target and good candidate in antiviral drug design and development research.

## **2.8 HCV NS5B**

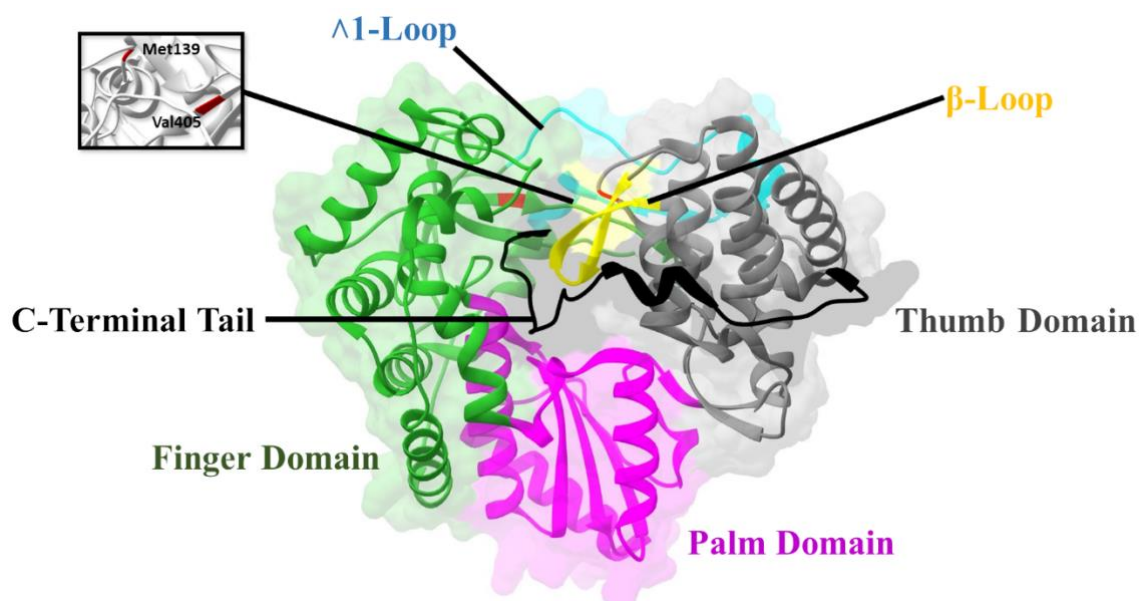
### *2.8.1 Structure, Function and Potential as a Drug Target*

The HCV RdRp is a 66 kDa protein, which is the key enzyme responsible for RNA synthesis and genome replication (Han *et al.*, 2014). The protein consists of 591 amino acids and is located at the C-terminal of the viral genome. Similar to other viral polymerases, the HCV RdRp adopts a right-handed topology and is subdivided into three domains: thumb (residues 371-561), finger (residues 1-187, 228-286) and palm (residues 188-227, 287-370) (Boyce *et al.*, 2014).

There are three important structural elements that are required for efficient viral replication:  $\Lambda$ 1-loop (residues 12-46),  $\beta$ -loop (residues 443-455) and C-terminal tail (residues 540-561) (Boyce *et al.*, 2014). The finger domain contains an extended loop known as the  $\Lambda$ 1-loop. The establishment of this interaction is imperative as it aids in communication of the two domains



during translocation of the RNA. The RdRp has two enzymatic activity modes parallel to differing conformations. Extensive interdomain interactions between the thumb and finger domains prompt the enzyme to take on a closed “active” conformation. This conformational state provides a suitable environment for *de novo* RNA synthesis for viral replication initiation (Sesmero and Thorpe, 2015). In 2002, Labonte *et al.*, showed the importance of the finger loop by conducting experimental studies which suggested that arginine or serine substitution of leucine 30 in the  $\Lambda$ 1-loop disrupts interactions between the thumb and finger domains and ultimately prevents the closed conformation from occurring, resulting in a non-functional enzyme (Labonté *et al.*, 2002). The C-terminal is made up of hydrophobic residues and forms the RdRp linker region, allowing protein membrane anchorage in cells, which is responsible for ER membrane association during viral replication. The C-terminal residues are in close proximity to another regulatory motif known as the  $\beta$ -loop. The  $\beta$ -loop is designated as a non-catalytic area that encompasses a regulatory motif, which expedites the inhibition of RdRp activity and prevents RNA binding. All three regulatory elements and domains are graphically represented in Figure 7.



**Figure 7:** The three-dimensional crystal structure of HCV NS5B polymerase (PDB ID: 3H5S) highlighting the three domains and important elements required for viral replication, generated using UCSF Chimera software program (Prepared by Author).

The RdRp is viewed as an attractive target for antiviral therapeutic intervention of HCV associated diseases due to its distinct and exclusive ability to use the RNA template that is absent in host mammalian cells (Scheel and Rice, 2013; Wei *et al.*, 2016). Currently, there are many X-ray crystallography structures of the free enzyme as well as inhibitor-bound HCV RdRp available,

which provide significant support for the discovery and development of new structure-based HCV RdRp inhibitors (Wei *et al.*, 2016).

## **2.9 HCV therapy**

### *2.9.1 Current Therapies*

Since the discovery of HCV, numerous therapeutics have been evaluated to detect effectiveness in the treatment of the chronic infection. A synthetic analogue of purine, known as Acyclovir, was the first drug to be clinically evaluated in HCV treatment. Regrettably, the drug showed no significant impact in treatment of the disease. Over there years, the medicinal and pharmaceutical domains have been able to identify successful and potential candidates in combined or mono-HCV therapy, these include ribavirin, interferon and direct acting antivirals that target viral proteins required for replication.

#### *2.9.1.1 Ribavirin and Pegylated Interferon-alpha*

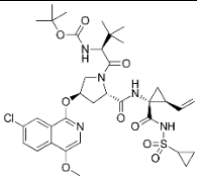
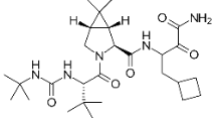
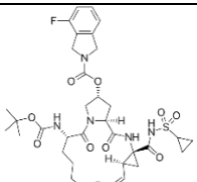
Ribavirin, a nucleoside inhibitor, is a ribonucleic-guanosine analogue that is used to impede synthesis of viral RNA and inhibit precursor messenger RNA capping. Metabolic processing of ribavirin gives rise to a molecule that bares close resemblance to purine nucleotides of RNA. This feature allows the molecule to disrupt viral RNA metabolism, which is a process necessary for efficient HCV replication. Unfortunately, ribavirin is ineffective against HCV on its own (Gish, 2006; Graci and Cameron, 2006). Interferons are naturally occurring proteins that form a vital part of the immune system. They function in two ways. Firstly, they hamper the process of viral replication and secondly, they augment immune responses. Interferons affect viral replication by binding to receptors presented by almost every cell type, stopping entry of the virus into the host cell as well as replication. These proteins are known to boost immune response by inducing immune cell activity, which in effect, renders the compromised cells more vulnerable to actions of the immune system (Hervas-Stubbs *et al.*, 2011). Interferon produced in laboratories are metabolised at an exceedingly rapid rate, as a result, its effectiveness is reduced, and the virus can replicate during the small intervals arising between treatments. To address this challenge, IFN was subjected to pegylation, a process by which a large molecular chain is attached to the protein to slow down the rate at which it is metabolised. Addition of the molecular chain allows consistent levels of the drug to be released into circulation, which projects a continual attack on the virus (Foster, 2010). However, many patients undergo PEGylated IFN- $\alpha$  therapy experienced debilitating adverse effects manifesting in the form of depression, headaches, shivering, fever and

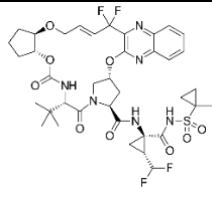
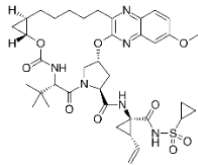
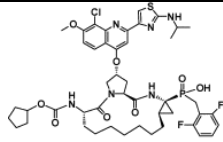
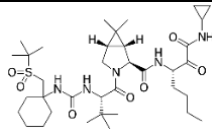
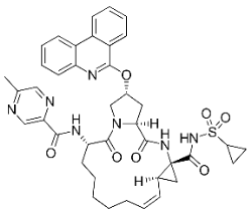
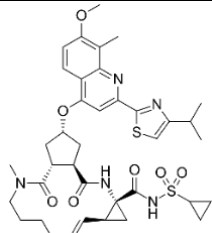
fatigue (Nguyen and Morgan, 2012). Together, ribavirin and PEGylated IFN- $\alpha$  form an effective combined therapeutic regime against chronic HCV infection eradicating the virus in 40-50% of HCV-infected individuals (Palumbo, 2011). Unfortunately, the ribavirin component of combined therapy has prevented many HCV-infected patients from successfully completing therapy due to the development of gout, sinusitis, rash, itching, fatigue and haemolytic anaemia. The use of ribavirin has also resulted in death by myocardial infarction in those presenting with cardiovascular disease (Koh and Liang, 2014).

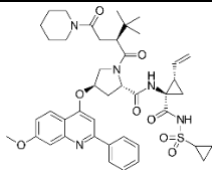
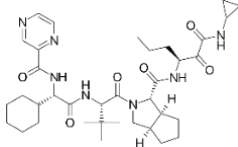
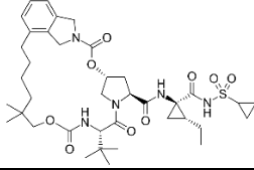
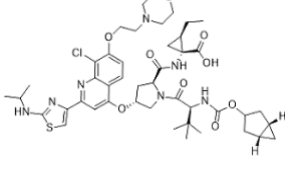
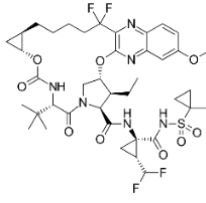
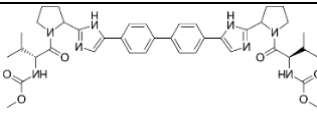
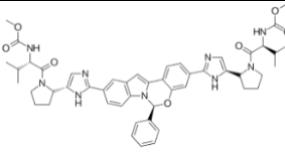
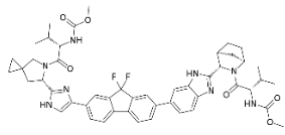
### 2.9.1.2 Direct-Acting Antiviral Drugs

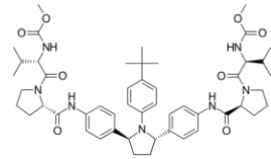
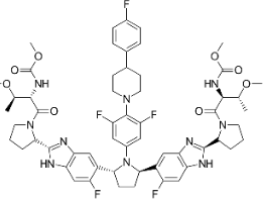
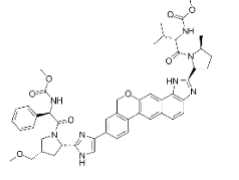
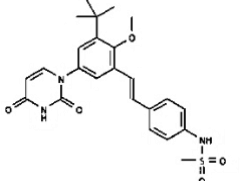
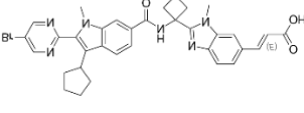
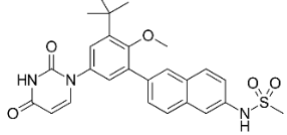
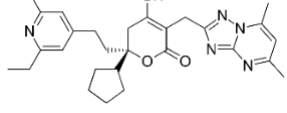
Direct-acting antiviral drugs (DAAs) are a new class of HCV drug. These drugs function by directly targeting specific stages in the viral life cycle of HCV. There are four classes of DAAs: NS3/4A protease inhibitors (PIs), Nucleotide/nucleoside NS5B RdRp inhibitors (NIs), NS5A inhibitors and lastly, non-nucleoside NS5B RdRp inhibitors (NNIs) (Pockros, 2015). The current drugs under investigation and/or Food and Drug Administration (FDA) approved drugs are represented in Table 1.

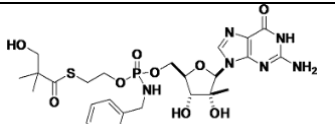
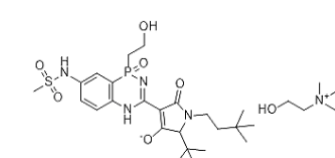
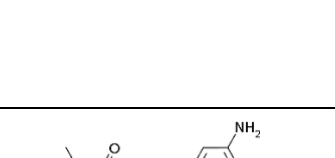
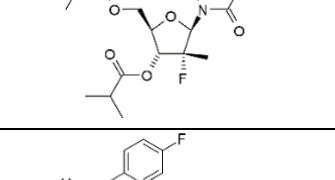
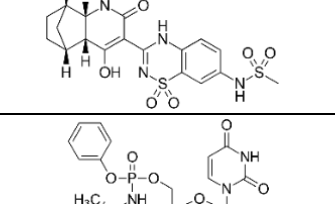
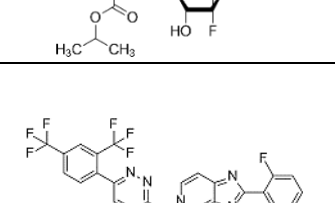
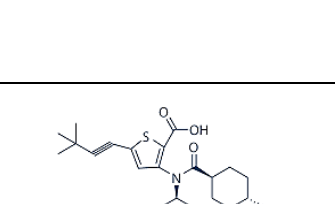
**Table 1:** Current HCV direct-acting antiviral drugs (DAAs) in Clinical Testing and/or FDA approved.

Drug	Structure	Phase	Class	Genotype
Asunaprevir		Phase II	PI	1, 4
Boceprevir		FDA approved Discontinued	PI	1
Danoprevir		Phase I	PI	1-6

Glecaprevir		FDA approved  Mavyret- glecaprevir, pibrentasvir	PI	1-6
Grazoprevir		FDA approved  Zepatier- grazoprevir, elbasvir with or without ribavirin	PI	1, 3, 4
GS-9256		Phase II	PI	1
Narlaprevir		Phase II	PI	1
Paritaprevir (ABT-450)		FDA approved  Viekira Pak- paritaprevir, ombitasvir, dasabuvir, and ritonavir  Technivie- paritaprevir, ombitasvir and ritonavir	PI	1, 4
Simeprevir		FDA approved  Olysio- simeprevir, ribavirin or sofosbuvir and PEG-IFN- $\alpha$	PI	1

Sovaprevir		Phase II	PI	1
Telaprevir		FDA approved	PI	1
Vaniprevir		Phase II	PI	1
Vedroprevir		Phase II	PI	1
Voxilaprevir		FDA approved Vosevi- Voxilaprevir, sofosbuvir and velpatasvir	PI	1-6
Daclatasvir		Daklinza- declatasvir, sofosbuvir, ribavirin, and interferon	NS5A	1-4
Elbasvir		FDA approved Zepatier- Elbasvir and grazoprevir	NS5A	1, 4
Ledipasvir		FDA approved Harvoni- ledipasvir and sofosbuvir	NS5A	1

Ombitasvir		FDA approved  Viekira Pak- Ombitasvir, paritaprevir, ritonavir	NS5A	1
Pibrentasvir		FDA approved  Mavyret- glecaprevir and pibrentasvir	NS5A	1-6
Velpatasvir		FDA approved  Epclusa- vepatasvir and sofosbuvir	NS5A	1-6
ABT-072		Phase II	NNI	1
Deleobuvir		Phase III Discontinued- insufficient effectiveness	NNI	1
Dasabuvir		FDA approved  Exviera- dasabuvir, ombitasvir, paritaprevir, ritonavir	NNI	1
Filibuvir		Phase II Discontinued- strategic reasons	NNI	1

IDX 184		Phase II	NI	1
IDX-375		Phase I Discontinued- Europe Phase II Discontinued- USA	NNI	1
Mericitabine		Phase II Discontinued	NI	1
Setrobuvir		Phase I	NNI	1
Sofosbuvir		FDA approved	NI	1-4
Tegobuvir		Phase II	NNI	1
Lomibuvir		Phase II	NNI	1

Numerous HCV antiviral drugs are currently undergoing clinical trials or in some cases have been discontinued due to poor pharmacodynamic profiles, as stated in Table 1. Only a few HCV drugs have been approved by the FDA for human administration. Even though these drugs are shown to be successful in the treatment of HCV infection, they are only effective in a small percentage

of patients. Most HCV-infected individuals do not adequately respond to current market-available antiviral drugs due to viral genotype, onset of severe side effects and the need of frequent administrations (Bertino *et al.*, 2016). The antiviral drugs are also shown to have a low barrier to viral resistance and lack selectivity for the virus and target mainly genotype 1 (Irwin *et al.*, 2016).

Advanced PIs, Boceprevir and Telaprevir and recently emerged, Narlaprevir, are reversible covalent inhibitors. Boceprevir has since been discontinued but Telaprevir in combination with ribavirin and INF- $\alpha$  are given in triple therapy to patients who were unsuccessful in previous therapies. There are, however, several complications with this HCV therapeutic regime. Firstly, Telaprevir is only effective when administered with PEGylated IFN- $\alpha$  and ribavirin caused by a low resistance barrier whereby single point mutations instigate resistance to the PI. Secondly, Telaprevir is only effective against specific genotypes of HCV. Thirdly, PIs are known to elicit side effects and lastly, triple therapy is expensive and is consequently a less attainable resource for patients with financial constraints. Therefore, new drugs are still required to make HCV treatment better tolerated and more accessible so it can tackle the global HCV burden (Hanson *et al.*, 2012).

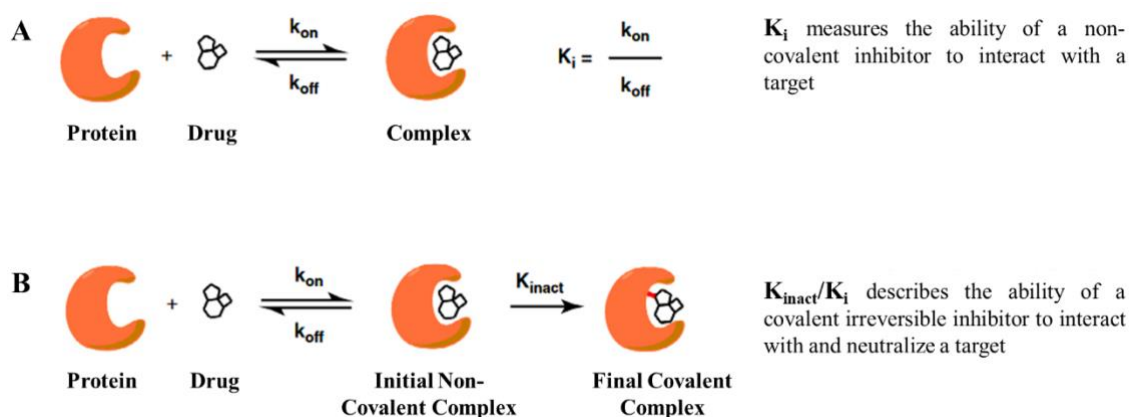
### 2.9.2 Drug Resistance

Viral resistance to HCV drugs has developed into an alarming problem that requires immediate attention. It is well known in the scientific community that viruses undergo constant modifications to allow cell survival and evasion of host defence mechanisms. Within humans, HCV can produce millions of new viruses daily. Furthermore, HCV RdRp lacks adequate proof-reading activities that precipitates at least 1 error amongst every 10 000 nucleotide bases generated (Sesmero and Thorpe, 2015). In essence, an infected individual can potentially carry more than just one quasispecies of HCV (Li and Lo, 2015). This genetic variability presents a difficult task in the design of anti-HCV drugs as a drug may be effective against a target, diminishing its viral load, only for evolutionary forces to select another resistant viral quasispecies to which the drug is ineffective and futile (Echeverria *et al.*, 2015). To combat this issue, a novel drug design approach was proposed, in this strategy, antiviral drugs were designed to specifically target a residue present within a biomolecular target. This strategy attained the term, selective covalent inhibition and involved irreversible binding of the inhibitor to the target protein.



## 2.10 Covalent Inhibition

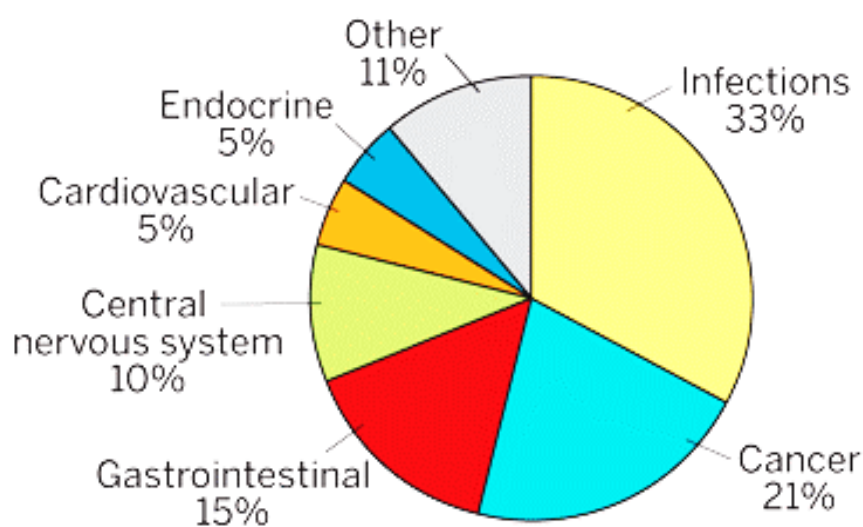
Majority of drug discovery endeavours have culminated in the form of non-covalent reversibly-bound drugs, although in recent years, mainstream attention has been directed toward the design and discovery of covalent inhibitors. Almost 30% of market-available drugs function through a covalent mechanism, although most of them were not deliberately produced as covalent inhibitors but rather discovered by chance (Singh *et al.*, 2011; Mah, Thomas and Shafer, 2014). These inhibitors undergo covalent binding to a designated residue of interest located in a target protein. Covalent inhibitors are rapidly emerging as a valuable therapeutic approach in antiviral drug discovery. Small inhibitor molecules are used as covalent inhibitors, which encompass a moiety that is explicitly designed to bind to an amino acid functional group of a target protein (Awoonor-Williams, Walsh and Rowley, 2017). This manifestation occurs through conventional reversible or irreversible interactions while enduring the formation of a persistent non-liable bond that strengthens the binding affinity between the target protein and covalent inhibitor (Becker *et al.*, 2016; Shunmugam, Ramharack and Soliman, 2017). The diagrammatic representation of conventional non-covalent and covalent irreversible inhibitor binding is depicted in Figure 8.



**Figure 8:** Schematic representation of conventional non-covalent (A) and covalent (B) drug interactions (Adapted from Bauer, 2015). In the first step (A) The drug binds to the target protein through a conventional non-covalent reversible inhibitory mechanism, which results in the formation of a protein-drug complex. (B) A non-covalent interaction initially occurs then a subsequent chemical reaction takes place between the protein and drug, in doing so, a covalent irreversibly bound complex is generated (Bauer, 2015; Shunmugam, Ramharack and Soliman, 2017).

Inhibitors covalently bind to their target protein, consequently disassociating from it at a rate greater than the turnover rate of protein synthesis (Baillie, 2016). This mechanism can be

observed in reactions between inhibitors and proteins carrying electrophilic “warheads” as active site electrophilic traps (Shunmugam, Ramharack and Soliman, 2017). These electrophilic warheads include aldehydes, activated ketones, alpha-ketoamides and keto-heterocycles, carbonitriles, and derivatives of boronic acid (Ghosh and Gemma, 2014; Baillie, 2016). Conversely, irreversible covalent inhibitors form adducts with their target protein which might not separate from the protein for the duration of its lifespan or do so with a kinetic half-life portentously longer than the rate of protein re-synthesis. Such compounds, frequently contain harder electrophilic warheads such as epoxides,  $\alpha$ -halo ketones, acyloxymethyl ketones, aziridines, vinyl sulfones, and activated acetylenes, granting most irreversible covalent inhibitors commonly engage Michael acceptors via the Michael addition reaction (Blake and Soliman, 2014; Baillie, 2016). This chemical reaction is extensively used to achieve irreversible covalent binding and the functional groups commonly involved include propargylic acid derivatives, alkynyl amides, quinones, vinyl sulfonates and acrylamides (Liu *et al.*, 2013). Currently there are approximately 39 covalent drugs available on the market and their associated distribution is graphically depicted in Figure 9 (Guterman, 2011).



**Figure 9:** Percentage distribution of covalent drugs currently used in the pharmaceutical industry (Guterman, 2011).

Numerous therapeutic benefits are expressed by covalent binding in comparison to conventional reversible binding, such as biochemical efficacy. Whereby non-equilibrium binding restricts the competition that may arise between elevated ligand or substrate concentrations, prolonged duration of drug action, requires less frequent doses, rapid target inactivation followed by prompt elimination, minimal drug to drug interaction, reduced occurrence of toxicities, lower drug dose

concentration is required, increased potency, diminished pharmacodynamic sensitivity and the ability to evade mechanisms of resistance (Lamarre *et al.*, 2003; Swinney, 2004; Kumalo, Bhakat and Soliman, 2015; Baillie, 2016). Figure 10 displays a list of FDA approved covalent inhibitors.

<b>Drug</b>	<b>Biological Target</b>	<b>Therapeutic Domain</b>
Amoxicillin	PBP	Anti-infective
Cefaclor/Ceclor	PBP	Anti-infective
Ceftriaxone/Rocephin	PBP	Anti-infective
Cefuroxime axetil/ceftin	PBP	Anti-infective
Cephalexin/keflex	PBP	Anti-infective
D-cycloserine/seromycin	Alanine racemase	Anti-infective
Fosfomycin/monurol	UDP-N-acetylglucosamine-3-enolpyruvyl-transferase	Anti-infective
Isoniazid	Enol-acyl carrier protein reductase	Anti-infective
Meropenem	PBP	Anti-infective
Omnicef	PBP	Anti-infective
Penicillin V	PBP	Anti-infective
Azacytidine	Methyltransferase	Cancer
Bortezomib	Proteasome	Cancer
Decitabine/azadC	Methyltransferase	Cancer
Dutasteride/avodart	5- $\alpha$ -Reductase	Cancer
Exemestane/Aromasin	Aromatase	Cardio-vascular
Floxuridine	Thymidylate synthase	Cardio-vascular
Gemcitabine/gemzar	Ribonucleoside reductatase	Cardio-vascular
Proscar/finasteride	5- $\alpha$ -Reductase	Cardio-vascular
Rasagiline	MAO-B	Parkinson's disease
Selegiline	MAO-B	Parkinson's disease
Warfarin	Vitamin K reductase	Cardio-vascular
Vigabatrin/sabril	GABA-Aminotransferase	Anti-epileptic
Nexium/esomeprazole	H+/K+ ATPase	Gastro-intestinal
Orlistat/	Lipase	Gastro-intestinal
Prevacid/lansoprazole	H+/K+ATPase	Gastro- intestinal
Prilosec/omeprazole	H+/K+ATPase	Gastro-intestinal
Protonix/pantoprazole	H+/K+ATPase	Gastro-intestinal
Aciphex/rabeprazol	H+/K+ATPase	Gastro-intestinal
Aspirin	Cyclooxygenase	Inflammation
Disulfiram/antabuse	Aldehyde dehydrogenase	Chronic alcoholism
Eflornithine	Ornithine decarboxylase	Hirsutism
Propylthiouracil/procasil	Thyroxine-5-deiodinase	Hyperthyroidism
Saxagliptin/Onglyza	DPP-IV	Anti-diabetic drug
Vildagliptin/Eugreas	DPP-IV	Anti-diabetic drug
Phenoxy-benzamine hydrochloride	$\alpha$ -Adrenoceptor	Cardio-vascular
mercaptopurine/purinthol	Purine-nucleotide synthesis	Cancer
Carbidopa/lodosyn	DOPA decarboxylase	CNS

**Figure 10:** List of FDA approved covalent inhibitors (Kumalo, Bhakat and Soliman, 2015).

Initially, the pharmaceutical industry was reluctant to pursue covalent inhibition due to major adjustments taking place in screening methods and concerns arising in safety. This include

idiosyncratic toxicity caused by target protein haptenization leading to immune system initiating host antibody response, reactivity will offset and prohibit any selectivity covalent modification of biomolecules. This will cause toxic injury leading to initiation of host cell damage response and immunogenicity caused by the formation of protein inhibitor adducts (Johnson, Weerapana and Cravatt, 2010; Johansson, 2012). Using rational drug design, scientist were able to address these issues which lead to the unexpected discovery of targeted viral inhibitors also referred to as selective covalent inhibitors (Singh, Petter and Kluge, 2010).

### *2.10.1 Selective Covalent Inhibition*

Most selective covalent inhibition approaches have been unearthed by targeting highly reactive cysteine thiolate at the catalytic or non-catalytic binding site of a target protein. The purpose of this strategy is to appropriately and successfully address issues arising in current antiviral therapies. In the process of selective covalent inhibition, the thiol functional group of a cysteine residue is equipped with enhanced reactive properties, which makes it a perfect candidate for engagement of a covalent bond. The ligand is able to align an electrophilic trap with a cysteine residue located in the target protein (Lagoutte, Patouret and Winssinger, 2017). The designated cysteine residue of the target protein must be conserved across all geno- and sub-types of the virus. This feature allows selective covalent inhibitors to effectively act against the virus regardless of geno- and sub-types as well as quasispecies that may be present within a host at a single given time. This feature will also be beneficial when targeting HCV on a broader, international scale. Thus a drug's ability to efficiently target a specific cysteine amino acid residue may provide further means of attaining selectivity to reduce the occurrence of drug resistance to antiviral therapeutics (Shunmugam, Ramharack and Soliman, 2017; Shunmugam and Soliman, 2018).

## References

Adinolfi, L. E., Restivo, L. and Marrone, A. (2013) 'The predictive value of steatosis in hepatitis C virus infection', *Expert Review of Gastroenterology and Hepatology*, 7(3), pp. 205–213. doi: 10.1586/egh.13.7.

Adler, B. (2014) *Current Topics in Microbiology and Immunology and Leptospirosis*. doi: 10.1007/978-3-662-45059-8.

Alter, M. J. (2007) 'Epidemiology of hepatitis C virus infection', *World Journal of Gastroenterology*, 13(17), pp. 2436–2441.

Ashfaq, U. A. *et al.* (2011) 'An overview of HCV molecular biology, replication and immune responses', *Virology Journal*. doi: 10.1186/1743-422X-8-161.

Awoonor-Williams, E., Walsh, A. G. and Rowley, C. N. (2017) 'Modeling Covalent-Modifier Drugs', *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, 1865(11), pp. 1–36. doi: <https://doi.org/10.1016/j.bbapap.2017.05.009>.

Axley, P. *et al.* (2018) 'Hepatitis C Virus and Hepatocellular Carcinoma: A Narrative Review', *Journal of Clinical and Translational Hepatology*, 6(1), pp. 79–84. doi: 10.14218/JCTH.2017.00067.

Baillie, T. A. (2016) 'Targeted Covalent Inhibitors for Drug Design', *Angewandte Chemie - International Edition*, pp. 13408–13421. doi: 10.1002/anie.201601091.

Bataller, R. and Brenner, D. . (2005) 'Liver fibrosis', *The Journal of Clinical Investigation*, 115(2), pp. 209–218. doi: 10.1172/JCI200524282.

Bauer, R. A. (2015) 'Covalent inhibitors in drug discovery: From accidental discoveries to avoided liabilities and designed therapies', *Drug Discovery Today*, pp. 1061–1073. doi: 10.1016/j.drudis.2015.05.005.

Becker, D. *et al.* (2016) 'Irreversible inhibitors of the 3C protease of Coxsackie virus through templated assembly of protein-binding fragments', *Nature Communications*, 7, p. 12761. doi: 10.1038/ncomms12761.

Bertino, G. *et al.* (2016) 'Chronic hepatitis C: This and the new era of treatment', *World Journal of*

*Hepatology*, 8(2), pp. 92–106. doi: 10.4254/wjh.v8.i2.92.

Blake, L. and Soliman, M. E. S. (2014) ‘Identification of irreversible protein splicing inhibitors as potential anti-TB drugs: Insight from hybrid non-covalent/covalent docking virtual screening and molecular dynamics simulations’, *Medicinal Chemistry Research*, 23(5), pp. 2312–2323. doi: 10.1007/s00044-013-0822-y.

Blight, K. J. (2011) ‘Charged Residues in Hepatitis C Virus NS4B Are Critical for Multiple NS4B Functions in RNA Replication’, *Journal of Virology*, p. JVI-00858. doi: 10.1128/JVI.00858-11.

Bose, S. K. and Ray, R. (2014) ‘Hepatitis C virus infection and insulin resistance’, *World Journal of Diabetes*, 5(1), p. 52.

Boyce, S. E. *et al.* (2014) ‘Structural and regulatory elements of HCV NS5B polymerase -  $\beta$ -loop and C-terminal tail - Are required for activity of allosteric thumb site II inhibitors’, *PLoS ONE*, 9(1), p. e84808. doi: 10.1371/journal.pone.0084808.

Brown, R. . *et al.* (2016) ‘Hepacivirus NS3/4A Proteases Interfere with MAVS Signaling in both Their Cognate Animal Hosts and Humans: Implications for Zoonotic Transmission’, *Journal of Virology*, p. JVI-01634.

Butler, B. *et al.* (2016) ‘Symptomatic Acute Hepatitis C Infection Following a Single Episode of Unprotected Sexual Intercourse’, *Case Reports in Infectious Diseases*, 2016. doi: 10.1155/2016/8639098.

Byrd, A. K. (2012) ‘Superfamily 2 helicases’, *Frontiers in Bioscience*, 17, pp. 2070–2088. doi: 10.2741/4038.

Carulli, L. and Anzivino, C. (2014) ‘Telomere and telomerase in chronic liver disease and hepatocarcinoma’, *World Journal of Gastroenterology*, 20(20), pp. 6287–6292. doi: 10.3748/wjg.v20.i20.6287.

Căruntu, F. A. and Benea, L. (2006) ‘Acute hepatitis C virus infection: Diagnosis, pathogenesis, treatment’, *Journal of Gastrointestinal and Liver Diseases*, 15(3), pp. 249–256. doi: 10.1021/np200993k.

Chandler, D. E. *et al.* (2012) ‘The p7 Protein of Hepatitis C Virus Forms Structurally Plastic, Minimalist

- Ion Channels', *PLoS Computational Biology*. doi: 10.1371/journal.pcbi.1002702.
- Chevaliez, S. and Pawlotsky, J.-M. (2006) 'HCV Genome and Life Cycle', *Hepatitis C Viruses: Genomes and Molecular Biology*, pp. 5–47. doi: PMID: 21250393.
- Choi, Y. H. *et al.* (2016) 'Elevation of alanine aminotransferase activity occurs after activation of the cell-death signaling initiated by pattern-recognition receptors but before activation of cytolytic effectors in NK or CD8+ T cells in the liver during acute HCV infection', *PLoS ONE*, 11(10), p. e0165533. doi: 10.1371/journal.pone.0165533.
- David, N. *et al.* (2015) 'The interaction between the Hepatitis C proteins NS4B and NS5A is involved in viral replication', *Virology*, 475, pp. 139–149. doi: 10.1016/j.virol.2014.10.021.
- Deutsch, M. *et al.* (2013) 'Clinical Characteristics, Spontaneous Clearance and Treatment Outcome of Acute Hepatitis C: A Single Tertiary Center Experience.', *Saudi Journal of Gastroenterology*, 19(2), pp. 81–85. doi: 10.4103/1319-3767.108479.
- Dixon, L. J. *et al.* (2013) 'Kupffer cells in the liver', *Comprehensive Physiology*, 3(2), pp. 785–797. doi: 10.1002/cphy.c120026.
- Donate, L. E. and Blasco, M. A. (2011) 'Telomeres in cancer and ageing', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1561), pp. 76–84. doi: 10.1098/rstb.2010.0291.
- Dustin, L. B. *et al.* (2016) 'Hepatitis C virus: life cycle in cells, infection and host response, and analysis of molecular markers influencing the outcome of infection and response to therapy', *Clinical Microbiology and Infection*. doi: 10.1016/j.cmi.2016.08.025.
- Echeverria, N. *et al.* (2015) 'Hepatitis C virus genetic variability and evolution', *World Journal of Hepatology*, 7(6), pp. 831–845. doi: 10.4254/wjh.v7.i6.831.
- Fedeli, U. *et al.* (2017) 'Mortality associated with hepatitis C and hepatitis B virus infection: A nationwide study on multiple causes of death data', *World Journal of Gastroenterology*, 23(10), pp. 1866–1871. doi: 10.3748/wjg.v23.i10.1866.
- Fonseca-Coronado, S. *et al.* (2012) 'Specific detection of naturally occurring hepatitis C virus mutants with resistance to telaprevir and boceprevir (protease inhibitors) among treatment-naïve infected individuals', *Journal of Clinical Microbiology*, 50(2), pp. 281–287. doi: 10.1128/JCM.05842-11.

Foster, G. R. (2010) 'Pegylated interferons for the treatment of chronic hepatitis C: Pharmacological and clinical differences between peginterferon- $\alpha$ -2a and peginterferon- $\alpha$ -2b', *Drugs*, 70(2), pp. 147–165. doi: 10.2165/11531990-000000000-00000.

Frick, D. N. (2006) 'HCV Helicase: Structure, Function, and Inhibition', in *Hepatitis C Viruses: Genomes and Molecular Biology*. Norfolk (UK): Horizon Bioscience.

Fujita, T. and Narumiya, S. (2016) 'Roles of hepatic stellate cells in liver inflammation: a new perspective', *Inflammation and Regeneration*, 36(1), p. 1. doi: 10.1186/s41232-016-0005-6.

Gawlik, K. and Gallay, P. A. (2014) 'HCV core protein and virus assembly: what we know without structures', *Immunologic Research*. doi: 10.1007/s12026-014-8494-3.

Gentsch, J. *et al.* (2013) 'Hepatitis C Virus p7 is Critical for Capsid Assembly and Envelopment', *PLoS Pathogens*. doi: 10.1371/journal.ppat.1003355.

Ghany, M. G. and Jake Liang, T. (2016) 'Natural history of chronic hepatitis C', in *Hepatitis C Virus II: Infection and Disease*. doi: 10.1007/978-4-431-56101-9\_1.

Ghosh, A. K. and Gemma, S. (2014) *NS3 / 4A Serine Protease Inhibitors for the Treatment of HCV: Design and Discovery of Boceprevir and Telaprevir*. First Edit. Wiley-VCH Verlag GmbH & Co. KGaA.

Gish, R. G. (2006) 'Treating HCV with ribavirin analogues and ribavirin-like molecules', *Journal of Antimicrobial Chemotherapy*, 57(1), pp. 8–13. doi: 10.1093/jac/dki405.

Golden-Mason, L. and Rosen, H. R. (2013) 'Natural killer cells: Multifaceted players with key roles in hepatitis C immunity', *Immunological Reviews*, 255(1), pp. 68–81. doi: 10.1111/imr.12090.

Goossens, N. and Hoshida, Y. (2015) 'Hepatitis C virus-induced hepatocellular carcinoma', *Clinical and Molecular Hepatology*, 21(2), pp. 105–114. doi: 10.3350/cmh.2015.21.2.105.

Gouttenoire, J. *et al.* (2014) 'Aminoterminal Amphipathic  $\alpha$ -Helix AH1 of Hepatitis C Virus Nonstructural Protein 4B Possesses a Dual Role in RNA Replication and Virus Production', *PLoS Pathogens*, 10(11), p. e1004501. doi: 10.1371/journal.ppat.1004501.

Gower, E. *et al.* (2014) 'Global epidemiology and genotype distribution of the hepatitis C virus



infection', *Journal of Hepatology*, 61(1), pp. S45–S57. doi: 10.1016/j.jhep.2014.07.027.

Graci, J. D. and Cameron, C. E. (2006) 'Mechanisms of action of ribavirin against distinct viruses', *Reviews in Medical Virology*, 16(1), pp. 37–48. doi: 10.1002/rmv.483.

Greener, M. (2013) 'Symptoms of liver disease', in *Coping with Liver Disease*.

Guterman, L. (2011) 'COVALENT DRUGS FORM LONG-LIVED TIES', *Chemical & Engineering News Archive*. doi: 10.1021/cen-v089n036.p019.

Han, S.-H. *et al.* (2014) 'Phosphorylation of Hepatitis C Virus RNA Polymerases Ser29 and Ser42 by Protein Kinase C-Related Kinase 2 Regulates Viral RNA Replication', *Journal of Virology*, p. JVI-01826. doi: 10.1128/JVI.01826-14.

Hanson, A. *et al.* (2012) 'Identification and Analysis of Inhibitors Targeting the Hepatitis C Virus NS3 Helicase', *Methods in Enzymology*, 511, pp. 463–483. doi: 10.1016/B978-0-12-396546-2.00021-8.

Hervas-Stubbs, S. *et al.* (2011) 'Direct effects of type I interferons on cells of the immune system', *Clinical Cancer Research*, 17(9), p. clincanres-1114. doi: 10.1158/1078-0432.CCR-10-1114.

Holder, K. A., Russell, R. S. and Grant, M. D. (2014) 'Natural Killer Cell Function and Dysfunction in Hepatitis C Virus Infection', *BioMed research international*, pp. 1–9. doi: 10.1155/2014/903764.

Hoshida, Y. *et al.* (2014) 'Pathogenesis and prevention of hepatitis C virus-induced hepatocellular carcinoma', *Journal of Hepatology*, 61(10), pp. S79–S90. doi: 10.1016/j.jhep.2014.07.010.

Hum, J. and Jou, J. H. (2018) 'The link between hepatitis C virus and diabetes mellitus: Improvement in insulin resistance after eradication of hepatitis C virus', *Clinical Liver Disease*, 11(3), pp. 73–76.

Irshad, M., Mankotia, D. S. and Irshad, K. (2013) 'An insight into the diagnosis and pathogenesis of hepatitis C virus infection', *World Journal of Gastroenterology*, 19(44), pp. 7896–7909. doi: 10.3748/wjg.v19.i44.7896.

Irwin, K. K. *et al.* (2016) 'Antiviral drug resistance as an adaptive process', *Virus Evolution*, 2(1), p. vew014. doi: 10.1093/ve/vew014.

Ivanov, A. V. *et al.* (2013) 'HCV and oxidative stress in the liver', *Viruses*, 5(2), pp. 439–469. doi:

10.3390/v5020439.

Johansson, M. H. (2012) 'Reversible Michael additions: covalent inhibitors and prodrugs', *Mini Rev Med Chem*, 12(13), pp. 1330–1344. doi: 10.2174/13895575112091330.

Johnson, D. S., Weerapana, E. and Cravatt, B. F. (2010) 'Strategies for discovering and derisking covalent, irreversible enzyme inhibitors', *Future Medicinal Chemistry*, 2(6), pp. 949–964. doi: 10.4155/fmc.10.21.

Jones, C. T. *et al.* (2007) 'Hepatitis C Virus p7 and NS2 Proteins Are Essential for Production of Infectious Virus', *Journal of Virology*. doi: 10.1128/JVI.00690-07.

Kalinina, O. V. and Dmitriev, A. V. (2015) 'Structural and functional genome organization and life cycle of hepatitis C virus', *Molecular Genetics, Microbiology and Virology*, 30(2), pp. 64–70. doi: 10.3103/S0891416815020044.

Karoney, M. J., Siika, A. M. and Karoney, M. J. (2013) 'Hepatitis C virus (HCV) infection in Africa: a review', *Pan African Medical Journal. Pan African Medical Journal*, 141444, pp. 44–1937. doi: 10.11604/pamj.2013.14.44.2199.

Kaushik, K. S., Kapila, K. and Praharaj, A. K. (2011) 'Shooting up: The interface of microbial infections and drug abuse', *Journal of Medical Microbiology*, 60(4), pp. 408–422. doi: 10.1099/jmm.0.027540-0.

Kazakov, T. *et al.* (2015) 'Hepatitis C Virus RNA Replication Depends on Specific Cis- and Trans-Acting Activities of Viral Nonstructural Proteins', *PLoS Pathogens*, 11(4), p. e1004817. doi: 10.1371/journal.ppat.1004817.

Khullar, V. and Firpi, R. J. (2015) 'Hepatitis C cirrhosis: New perspectives for diagnosis and treatment', *World Journal of Hepatology*, 7(14), pp. 1843–1855. doi: 10.4254/wjh.v7.i14.1843.

Kim, C. W. and Chang, K.-M. (2013) 'Hepatitis C virus: virology and life cycle', *Clinical and Molecular Hepatology*, 19(1), pp. 17–25. doi: 10.3350/cmh.2013.19.1.17.

Kim, S., Han, K. . and Ahn, S. . (2016) 'Hepatitis C Virus and Antiviral Drug Resistance', *Gut and liver*, 10(6), p. 890. doi: 10.5009/gnl15573.

Koh, C. and Liang, T. J. (2014) 'What is the future of ribavirin therapy for hepatitis C?', *Antiviral*

*Research*, 104, pp. 34–39. doi: 10.1016/j.antiviral.2014.01.005.

Kralji, D. *et al.* (2016) ‘Hepatitis C Virus, Insulin Resistance, and Steatosis’, *Journal of Clinical and Translational Hepatology*, 4(1), pp. 66–75. doi: 10.14218/JCTH.2015.00051.

Kumalo, H. M., Bhakat, S. and Soliman, M. E. S. (2015) ‘Theory and applications of covalent docking in drug discovery: Merits and pitfalls’, *Molecules*, 20(2), pp. 1984–2000. doi: 10.3390/molecules20021984.

Kumthip, K. and Maneekarn, N. (2015) ‘The role of HCV proteins on treatment outcomes’, *Virology Journal*. *Virology Journal*, 12(1), pp. 1–12. doi: 10.1186/s12985-015-0450-x.

Labonté, P. *et al.* (2002) ‘Modulation of hepatitis C virus RNA-dependent RNA polymerase activity by structure-based site-directed mutagenesis’, *Journal of Biological Chemistry*, 277, pp. 38838–38846. doi: 10.1074/jbc.M204657200.

Lagoutte, R., Patouret, R. and Winssinger, N. (2017) ‘Covalent inhibitors: an opportunity for rational target selectivity’, *Current Opinion in Chemical Biology*, 39, pp. 54–63. doi: 10.1016/j.cbpa.2017.05.008.

Lam, A. M. I., Keeney, D. and Frick, D. N. (2003) ‘Two Novel Conserved Motifs in the Hepatitis C Virus NS3 Protein Critical for Helicase Action’, *Journal of Biological Chemistry*, 278(48), pp. 44514–44524. doi: 10.1074/jbc.M306444200.

Lamarre, D. *et al.* (2003) ‘An NS3 protease inhibitor with antiviral effects in humans infected with hepatitis C virus’, *Nature*, 426(6963). doi: 10.1038/nature02099.

Lavanchy, D. (2011) ‘Evolving epidemiology of hepatitis C virus’, *Clinical Microbiology and Infection*. doi: 10.1111/j.1469-0691.2010.03432.x.

Lavie, M., Goffard, A. and Dubuisson, J. (2007) ‘Assembly of a functional HCV glycoprotein heterodimer’, *Current Issues in Molecular Biology*. doi: NBK1628 [bookaccession].

Law, J. L. M. *et al.* (2018) ‘Role of the E2 hypervariable region (HVR1) in the immunogenicity of a recombinant HCV vaccine’, *Journal of Virology*, 92(11), p. JVI.02141-17. doi: 10.1128/JVI.02141-17.

Li, H. C. and Lo, S. Y. (2015) ‘Hepatitis C virus: Virology, diagnosis and treatment’, *World Journal of*

*Hepatology*, 7(10), pp. 1377–1389. doi: 10.4254/wjh.v7.i10.1377.

Li, K. and Lemon, S. M. (2013) ‘Innate immune responses in hepatitis C virus infection’, *Seminars in Immunopathology*, 35(1), pp. 53–72. doi: 10.1007/s00281-012-0332-x.

Liang, Y. *et al.* (2018) ‘Hepatitis C virus NS4B induces the degradation of TRIF to inhibit TLR3-mediated interferon signaling pathway’, *PLoS Pathogens*, 14(5), p. e1007075. doi: 10.1371/journal.ppat.1007075.

Liu, Q. *et al.* (2013) ‘Developing irreversible inhibitors of the protein kinase cysteinome’, *Chemistry and Biology*, 20(2), pp. 146–159. doi: 10.1016/j.chembiol.2012.12.006.

Lohmann, V. *et al.* (1999) ‘Replication of subgenomic hepatitis C virus RNAs in a hepatoma cell line’, *Science*. doi: 10.1126/science.285.5424.110.

Ma, Y. *et al.* (2011) ‘Hepatitis C Virus NS2 Protein Serves as a Scaffold for Virus Assembly by Interacting with both Structural and Nonstructural Proteins’, *Journal of Virology*. doi: 10.1128/JVI.01070-10.

Madan, V. and Bartenschlager, R. (2015) ‘Structural and functional properties of the hepatitis C virus p7 viroporin’, *Viruses*. doi: 10.3390/v7082826.

Mah, R., Thomas, J. R. and Shafer, C. M. (2014) ‘Drug discovery considerations in the development of covalent inhibitors’, *Bioorganic and Medicinal Chemistry Letters*, 24(1), pp. 33–39. doi: 10.1016/j.bmcl.2013.10.003.

Masaki, T. *et al.* (2014) ‘Involvement of Hepatitis C Virus NS5A Hyperphosphorylation Mediated by Casein Kinase I- in Infectious Virus Production’, *Journal of Virology*, 88(13), pp. 7541–7555. doi: 10.1128/JVI.03170-13.

Masaki, T. and Suzuki, T. (2015) ‘NS5A phosphorylation: its functional role in the life cycle of hepatitis C virus’, *Future Virology*, 10(6), pp. 751–762. doi: 10.2217/fvl.15.33.

McGivern, D. R. *et al.* (2015) ‘Protease Inhibitors Block Multiple Functions of the NS3/4A Protease-Helicase during the Hepatitis C Virus Life Cycle’, *Journal of Virology*. doi: 10.1128/JVI.03188-14.

Medvedev, R., Ploen, D. and Hildt, E. (2016) ‘HCV and Oxidative Stress: Implications for HCV Life

Cycle and HCV-Associated Pathogenesis’, *Oxidative Medicine and Cellular Longevity*. doi: 10.1155/2016/9012580.

Mohamed, A. A. *et al.* (2015) ‘Hepatitis C virus: A global view’, *World Journal of Hepatology*. doi: 10.4254/wjh.v7.i26.2676.

Moosavy, S. H. *et al.* (2017) ‘Epidemiology, transmission, diagnosis, and outcome of Hepatitis C virus infection’, *Electronic Physician*, 9(10), p. 5646. doi: 10.19082/5646.

Moradpour, D. and Penin, F. (2013) ‘Hepatitis C virus proteins: From structure to function’, in *Hepatitis C virus: from molecular virology to antiviral therapy*. Springer, Berlin, Heidelberg, pp. 113–142. doi: 10.1007/978-3-642-27340-7-5.

Nault, J. C. *et al.* (2014) ‘Telomerase reverse transcriptase promoter mutation is an early somatic genetic alteration in the transformation of premalignant nodules in hepatocellular carcinoma on cirrhosis’, *Hepatology*, 60(6), pp. 1983–1992. doi: 10.1002/hep.27372.

Nguyen, D. L. and Morgan, T. R. (2012) ‘Management of Adverse Events During the Treatment of Chronic Hepatitis C Infection’, *Clinical Liver Disease*, 1(2), pp. 54–57. doi: 10.1002/cld.33.

Palomares-Jerez, F., Nemesio, H. and Villalaín, J. (2012) ‘The membrane spanning domains of protein NS4B from hepatitis C virus’, *Biochimica et Biophysica Acta - Biomembranes*, 1818(12), pp. 2958–2966. doi: 10.1016/j.bbamem.2012.07.022.

Palumbo, E. (2011) ‘Pegylated interferon and ribavirin treatment for hepatitis C virus infection’, *Therapeutic Advances in Chronic Disease*, 2(1), pp. 39–45. doi: 10.1177/2040622310384308.

Park, S. B. *et al.* (2016) ‘Hepatitis C Virus Frameshift/Alternate Reading Frame Protein Suppresses Interferon Responses Mediated by Pattern Recognition Receptor Retinoic-Acid-Inducible Gene-I’, *Plos One*, 11(7), p. e0158419. doi: 10.1371/journal.pone.0158419.

Park, S. H. and Rehermann, B. (2014) ‘Immune responses to HCV and other hepatitis viruses’, *Immunity*, 40(1), pp. 13–24. doi: 10.1016/j.immuni.2013.12.010.

Patel, A. and Harrison, S. A. (2012) ‘Hepatitis C virus infection and nonalcoholic steatohepatitis’, *Gastroenterology and Hepatology*, 8(5), pp. 305–312.

Paul, D. *et al.* (2011) 'NS4B Self-Interaction through Conserved C-Terminal Elements Is Required for the Establishment of Functional Hepatitis C Virus Replication Complexes', *Journal of Virology*, p. JVI-00502. doi: 10.1128/JVI.00502-11.

Pawlotsky, J.-M. (2013) 'NS5A inhibitors in the treatment of hepatitis C.', *Journal of hepatology*, 59(2), pp. 375–82. doi: 10.1016/j.jhep.2013.03.030.

Petruzzello, A. *et al.* (2016) 'Global epidemiology of hepatitis C virus infection: An up-date of the distribution and circulation of hepatitis C virus genotypes', *World Journal of Gastroenterology*, 22(34), pp. 7824–7840. doi: 10.3748/wjg.v22.i34.7824.

Phaniendra, A., Jestadi, D. B. and Periyasamy, L. (2015) 'Free Radicals: Properties, Sources, Targets, and Their Implication in Various Diseases', *Indian Journal of Clinical Biochemistry*, 30(1), pp. 11–26. doi: 10.1007/s12291-014-0446-0.

Plesa, G. *et al.* (2006) 'Immunogenicity of Cytopathic and Noncytopathic Viral Vectors', *Journal of Virology*, 80(13), pp. 6259–6299. doi: 10.1128/JVI.00084-06.

Pockros, P. J. (2015) 'Direct-acting antivirals for the treatment of hepatitis C virus infection', *UpToDate*. Accessed February, 28.

Pozzetto, B. *et al.* (2014) 'Health care-associated hepatitis C virus infection', *World Journal of Gastroenterology*, 20(46), p. 17265. doi: 10.3748/wjg.v20.i46.17265.

Rahal, A. *et al.* (2014) 'Oxidative stress, prooxidants, and antioxidants: The interplay', *BioMed Research International*, p. 761264. doi: 10.1155/2014/761264.

Rimmert, B. *et al.* (2014) 'A 3D structural model and dynamics of hepatitis C virus NS3/4A protease (genotype 4a, strain ED43) suggest conformational instability of the catalytic triad: implications in catalysis and drug resistivity', *Journal of Biomolecular Structure and Dynamics*, 32(6), pp. 950–958.

Salam, K. A. and Akimitsu, N. (2013) 'Hepatitis C Virus NS3 Inhibitors: Current and Future Perspectives', *BioMed research international*, pp. 1–9. doi: 10.1155/2013/467869.

Scheel, T. K. H. and Rice, C. M. (2013) 'Understanding the hepatitis C virus life cycle paves the way for highly effective therapies', *Nature Medicine*, 19(7), pp. 837–849. doi: 10.1038/nm.3248.

Sebastiani, G., Gkouvatsos, K. and Pantopoulos, K. (2014) 'Chronic hepatitis C and liver fibrosis', *World Journal of Gastroenterology*, 20(32), pp. 11033–11053. doi: 10.3748/wjg.v20.i32.11033.

Sesmero, E. and Thorpe, I. F. (2015) 'Using the hepatitis C virus RNA-dependent RNA polymerase as a model to understand viral polymerase structure, function and dynamics', *Viruses*, 7(7), pp. 3974–3994. doi: 10.3390/v7072808.

Shiffman, M. L. (2012) *Chronic hepatitis C virus: Advances in treatment, promise for the future*, *Chronic Hepatitis C Virus: Advances in Treatment, Promise for the Future*. doi: 10.1007/978-1-4614-1192-5.

Shiryaev, S. A. *et al.* (2012) 'New details of HCV NS3/4A proteinase functionality revealed by a high-throughput cleavage assay', *PLoS ONE*. doi: 10.1371/journal.pone.0035759.

Shunmugam, L., Ramharack, P. and Soliman, M. E. S. (2017) 'Road Map for the Structure-Based Design of Selective Covalent HCV NS3/4A Protease Inhibitors', *Protein Journal*. Springer US, 36(5), pp. 397–406. doi: 10.1007/s10930-017-9736-8.

Shunmugam, L. and Soliman, M. E. S. (2018) 'Targeting HCV polymerase: a structural and dynamic perspective into the mechanism of selective covalent inhibition', *RSC Advances*, 8(73), pp. 42210–42222.

Silini, E. *et al.* (1996) 'Hepatitis C virus genotypes and risk of hepatocellular carcinoma in cirrhosis: A case-control study', *Gastroenterology*, 111(1), pp. 199–205. doi: 10.1053/gast.1996.v111.pm8698200.

Singh, J. *et al.* (2011) 'The resurgence of covalent drugs.', *Nature reviews. Drug discovery*, 10(4), pp. 307–317. doi: 10.1038/nrd3410.

Singh, J., Petter, R. C. and Kluge, A. F. (2010) 'Targeted covalent drugs of the kinase family', *Current Opinion in Chemical Biology*, 14(4), pp. 475–480. doi: 10.1016/j.cbpa.2010.06.168.

Stambouli, O. (2014) *Hepatitis C Virus: Molecular Pathways and Treatments*.

Stapleford, K. A. and Lindenbach, B. D. (2011) 'Hepatitis C Virus NS2 Coordinates Virus Particle Assembly through Physical Interactions with the E1-E2 Glycoprotein and NS3-NS4A Enzyme Complexes', *Journal of Virology*. doi: 10.1128/JVI.02268-10.

Swinney, D. C. (2004) 'Opinion: Biochemical mechanisms of drug action: what does it take for

success?’, *Nature Reviews Drug Discovery*, 3(9), pp. 801–808. doi: 10.1038/nrd1500.

Takamizawa, A. *et al.* (1991) ‘Structure and organization of the hepatitis C virus genome isolated from human carriers’, *Journal of Virology*. doi: 10.1016/j.jpcs.2011.01.005.

Thrift, A. P., El-Serag, H. B. and Kanwal, F. (2017) ‘Global epidemiology and burden of HCV infection and HCV-related disease’, *Nature Reviews Gastroenterology and Hepatology*. doi: 10.1038/nrgastro.2016.176.

Thursz, M. and Fontanet, A. (2014) ‘HCV transmission in industrialized countries and resource-constrained areas’, *Nature Reviews Gastroenterology and Hepatology*, 11(1), p. 28. doi: 10.1038/nrgastro.2013.179.

Toshikuni, N., Arisawa, T. and Tsutsumi, M. (2014) ‘Hepatitis C-related liver cirrhosis - strategies for the prevention of hepatic decompensation, hepatocarcinogenesis, and mortality’, *World Journal of Gastroenterology*, 20(11), pp. 2876–2887. doi: 10.3748/wjg.v20.i11.2876.

Tosone, G. *et al.* (2014) ‘Vertical hepatitis C virus transmission: Main questions and answers’, *World Journal of Hepatology*, 6(8), p. 538. doi: 10.4254/wjh.v6.i8.538.

Vieyres, G., Dubuisson, J. and Pietschmann, T. (2014) ‘Incorporation of hepatitis C virus E1 and E2 glycoproteins: The keystones on a peculiar virion’, *Viruses*. doi: 10.3390/v6031149.

Villani, R., Bellanti, F. and Serviddio, G. (2013) ‘Treatment of Chronic HCV Infection in the Era of Protease Inhibitors’, in *Practical Management of Chronic Viral Hepatitis*, pp. 167–184.

Wang, Y. *et al.* (2013) ‘Hepatic stellate cells, liver innate immunity, and hepatitis C virus’, *Journal of Gastroenterology and Hepatology (Australia)*, 28(1), pp. 112–115. doi: 10.1111/jgh.12023.

Wang, Y. *et al.* (2017) ‘The unexpected structures of hepatitis C virus envelope proteins’, *Experimental and Therapeutic Medicine*, 14(3), pp. 1859–1865. doi: 10.3892/etm.2017.4745.

Wei, Y. *et al.* (2016) ‘Discovery of Novel Hepatitis C Virus NS5B Polymerase Inhibitors by Combining Random Forest, Multiple e-Pharmacophore Modeling and Docking’, *PLoS ONE*, 11(2), p. e0148181. doi: 10.1371/journal.pone.0148181.

World Health Organization (2018) *Fifth meeting of the Emergency Committee under the International*



*Health Regulations (2005) regarding microcephaly, other neurological disorders and Zika virus.*

Wuytack, F. *et al.* (2018) 'Sexual transmission of Hepatitis C Virus infection in a heterosexual population: A systematic review', *HRB Open Research*, 1.

Yan, Z. and Wang, Y. (2017) 'Viral and host factors associated with outcomes of hepatitis C virus infection (Review)', *Molecular medicine reports*, 15(5), pp. 2909–2924. doi: 10.3892/mmr.2017.6351.

Yu, G. Y. *et al.* (2012) 'Hepatic expression of HCV RNA-dependent RNA polymerase triggers innate immune signaling and cytokine production', *Molecular Cell*, 48(2), pp. 313–321. doi: 10.1016/j.molcel.2012.07.032.

## CHAPTER 3

### 3. Molecular Modelling and Computational Approaches to Biomolecular Structure and Covalent Drug Design

#### 3.1 Introduction

Molecular modelling is one of the most rapidly growing fields in science and is considered a fundamental component of physio-chemical and biological research. It is comprised of numerous indispensable computational and theoretical tools that are combined with the principles of chemistry, to model and mimic biomolecular systems. By doing so, researchers are able to understand the atomistic and molecular behaviour of these systems, from a realistic perspective (Kore *et al.*, 2012). In the last ten years, the rising expansion and subsequent popularity of molecular modelling is mainly owed to the substantial increase in the availability of protein structural data generated from nuclear magnetic resonance (NMR) spectroscopy, advanced electron microscopy and X-ray crystallography studies (Gorrec, 2014; Katebi *et al.*, 2015). There has also been revolutionary developments in the visualisation of molecular models due to vast enhancements in computer hardware and software, which may allow the handling of more complex systems (Goldbeck, 2012; Schlick, 2013; Kozlíková *et al.*, 2017). Although experimental methods can determine a biological system's mechanism of action, the necessity of its application falls short due to the intensive labour, extended time periods and high-cost requirements. For these reasons, researchers in the scientific community have governed the use of computational alternatives.

Drug design and discovery is a vibrant subject, cemented by its benefits of time efficient and cost-effective means to manufacture and produce new compound molecules. Over the years, several computational approaches have positively impacted the field of computer-aided drug design (CADD) (Baldi, 2010). These approaches are divided into three categories: (1) conformational modelling- associated with small molecule and macromolecule complexes, (2) property modelling- involving physical, chemical and biological properties and lastly (3) molecular design- which facilitates the optimisation of these properties (Mohan *et al.*, 2012). Thus, it can be asserted that computational approaches provide the pharmaceutical industry with a powerful tool, to accelerate drug design research for the discovery of novel drug targets as well as promoting the development of new promising drugs (Baldi, 2010; Lindahl, 2015).

Proceeding its conceptualisation in the 1960s, computational approaches have equipped the pharmaceutical industry and associated disciplines with a robust platform for drug discovery and *in silico* analysis of molecular structures (Baldi, 2010; Sliwoski *et al.*, 2014). Rational drug design (RDD) adheres to a three-step process, target identification, full characterisation of target, and design of potential molecules that binds to the target (Todd, Anderson and Groundwater, 2009). The notion of RDD was founded on the basic knowledge that activities displayed by drugs are gained following compound binding to a protein pocket within a biological target. Thereafter, the conformational and chemical stability at the binding site shown by the drug directly correlates to its potential as a successful candidate (Stocks, 2013). The evolutionary success of RDD is significantly supported by the materialisation of advancements in protein crystallography, computational chemistry and molecular biology. The majority of the approaches employed in RDD were initially developed for non-covalent drugs; though, these methods can be functionally adapted to suit covalent drugs. However an added necessity for specialised computational software and tools is to establish the specialised bond between the ligand and amino acid substrate (Lonsdale and Ward, 2018).

Structural analysis of biomolecules and RDD use numerous computational techniques including: molecular docking, virtual screening, multiple sequence analysis, homology modelling, covalent docking and molecular dynamic simulations. These techniques are widely used to understand a drug's binding mechanism, from a molecular perspective. There are two major molecular modelling principles that are used to determine energetics and conformational modifications of drug-protein systems:

- (1) Quantum Mechanics and
- (2) Molecular Mechanics

By combining these two principles with molecular dynamic simulations, the flexibility of the target protein and the significance of inhibitor binding may be effectively studied.

In this chapter, quantum mechanics, molecular mechanics and molecular dynamic simulations will be explained, thus providing comprehensive insight into the rationale underlying the chosen energy descriptors of this study. The principle behind the individual computational tools utilised in the study will also be further described.

### **3.2 The principle of Quantum Mechanics**

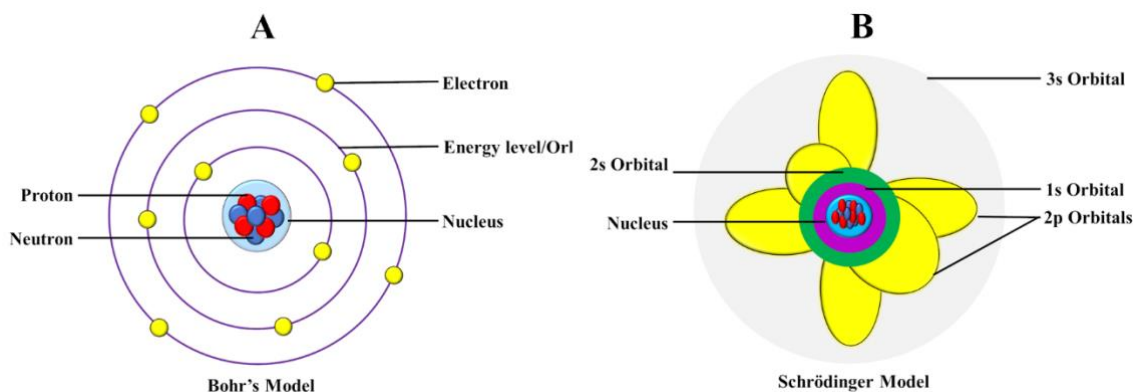
Quantum mechanics (QM) is one of the most successful branches of classical physics. The principles involved in QM were first established in the early 20<sup>th</sup> century and were categorised into two concepts. The first concept was conceived and developed by Planck, Born, Jordan, Heisenberg and Dirac, this theory was known as matrix mechanics. The second concept encompasses wave mechanics, which was developed by Erwin Schrodinger, to assist in the comprehension of quantum behaviour at a macroscopic level rather than the atomistic level (Tatulian, 2018).

The theory underlying QM explains characteristic behaviours of subatomic constituents that make up all forms of matter such as photons, neutrons, protons and electrons, at a nanoscopic scale (Jensen, 2017). Quantum theory plays a pivotal role in biological processes within the field of molecular biology such as electron excitation, atomic transfer, bond formation and breakage. In theory, QM has the ability to predict any individual property of a system, within a 3D space. Mapping of electrons is carried out by the continuous electron density method and system energy calculations are assessed using Schrödinger's wave function theory. When dealing with larger systems, electron density calculation may require the Born-Oppenheimer approximation theory.

Provided below is the basic principle of Schrödinger's wave function and Born-Oppenheimer approximation theory:

#### ***3.2.1 The Schrödinger Wave Function***

Originally, in 1913, Niels Bohr developed an atom model, proposing that electrons are arranged and revolve around in fixed concentric circular orbits surrounding the nucleus, like planets around the sun (Figure 1 A) (Bohr, 1913). In 1926, Erwin Schrödinger, an Austrian physicist took the Bohr atom model one step further, by using mathematical equations for probability (Schrödinger, 1926). This atom model was known as the quantum mechanical atom model and it suggested that one could only describe the probability of where an electron could be rather than define its precise path (Figure 1 B). Furthermore, Schrödinger also expressed that electrons were in a constant state of motion and didn't have one fixed or definite position in atoms, as previously claimed by Bohr. (Tatulian, 2018).



**Figure 1:** Schematic representation of (A) Niels Bohr and (B) Erwin Schrödinger atom models (Prepared by Author).

Through Schrödinger’s atom model and calculations, it was discovered that contrary to beliefs, electrons do not move in fixed circular orbits. Instead, electrons are in spatial regions called “clouds”. The first electron cloud is of the lowest energy level, denoted as 1s orbital. When an electron gains energy, it will jump to the next cloud, designated as the 2s orbital. From there, the electron will move to 2p, 3s/3p and so on. The clouds contain electrons of both low and high densities (Wang, 2016). In accordance with QM, all particles are described as a wave function with no momentum or distinct locality, until they are observed. The probability of each potential observation may be discerned by the wave function (Atkins and Friedman, 2011).

The fundamental basis of QM is built on the Schrödinger equation, in which the addition of atomic charge and mass allows prediction of nuclei and electron motions within a biomolecular system (Schlick, 2013; Bahrami *et al.*, 2014).

The Schrödinger wave equation can be represented as follows:

$$\mathbf{H}\psi = \mathbf{E} \psi \quad (\text{Eq 3.2.1})$$

Where: H denotes the Hamiltonian operator that encompasses derivatives with respect to the location of the atom,  $\psi$  is the wave function and E represents the system’s energy eigenvalues. In order to create a replicate of a physical model of Schrödinger’s equation, it is highly necessary for the wave function to be antisymmetric, normalised, mono-value and continuous. The

molecular Hamiltonian operator is calculated as the sum of kinetic energy (T) and the atom's total potential energy (V):

$$\mathbf{H} = \mathbf{T} + \mathbf{V} \quad (\text{Eq 3.2.2})$$

Where  $H$  is defined as follows:

$$\mathbf{H} = \left[ -\frac{\hbar^2}{8\pi^2} \sum_i \frac{1}{m_j} \left( \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) \right] + \sum_i \sum_{<j} \left( \frac{e_i e_j}{r_{ij}} \right) \quad (\text{Eq 3.2.3})$$

As observed, the Schrödinger equation is a highly complex calculation, thus its execution may be difficult to near impossible when dealing with molecular systems. This is due to molecular systems containing an excessively large number of atoms, as a consequence, the Born-Oppenheimer Approximation is used, This QM theory compensates for the molecular structure rather than the atomic structure (Schlick, 2013; Bahrami *et al.*, 2014).

### 3.2.2 The Born-Oppenheimer Approximation Theory

The Born-Oppenheimer approximation is one of the fundamental concepts underlying the description of molecular quantum mechanics. Developed in 1927 by Max Born and Robert Oppenheimer, this approximation is based on the assumption that the motion of the nuclei can be separated from electron motion (Born and Oppenheimer, 1927; Schlick, 2013). This theory is derived from the consideration of atomic particle mass and their subsequent influence they exert on each other. Interactions arise between the nucleus and electron, which represent the nature of the atom. Electrons exist at a much lighter mass in comparison to an atomic nucleus, due to their size, electrons have increased velocity and move almost instantaneously whereas the nuclei position is assumed to be fixed due to its heavier stature (Schlick, 2013). Since the nucleus is heavier, it is assumed to remain at a fixed position whilst dominating and steering the electron. Consequently, this causes the rapid oscillation of electrons around the nucleus through nuclear interactive forces. Hence, the Born-Oppenheimer approximation neglects the atomic nuclei motion when describing atomic electrons (Scherrer *et al.*, 2017). This permits the resolution of the Schrödinger equation regarding the kinetic energy of electrons; the nuclei kinetic energy remains constant (Sutcliffe and Woolley, 2012).

The velocity differences arising between the electrons and nuclei allow for the Born-Oppenheimer approximation to be applied, thus minimising the complexity of the wave function of the

Hamiltonian equation (Doltsinis, 2006; Malhado, Bearpark and Hynes, 2014). The simplified wave function is represented as follows:

$$\Psi(\mathbf{r}_{\text{elec}}) = \Psi(\mathbf{r}_{\text{elec}})(\Psi(\mathbf{r}_{\text{nuc}})) \quad (\text{Eq 3.2.4})$$

Equation 3.2.1 is converted to:

$$\mathbf{H}_{\text{EN}}\Psi(\mathbf{r}_{\text{elec}}) = \mathbf{E}_{\text{EN}}\Psi(\mathbf{r}_{\text{elec}}) \quad (\text{Eq 3.2.5})$$

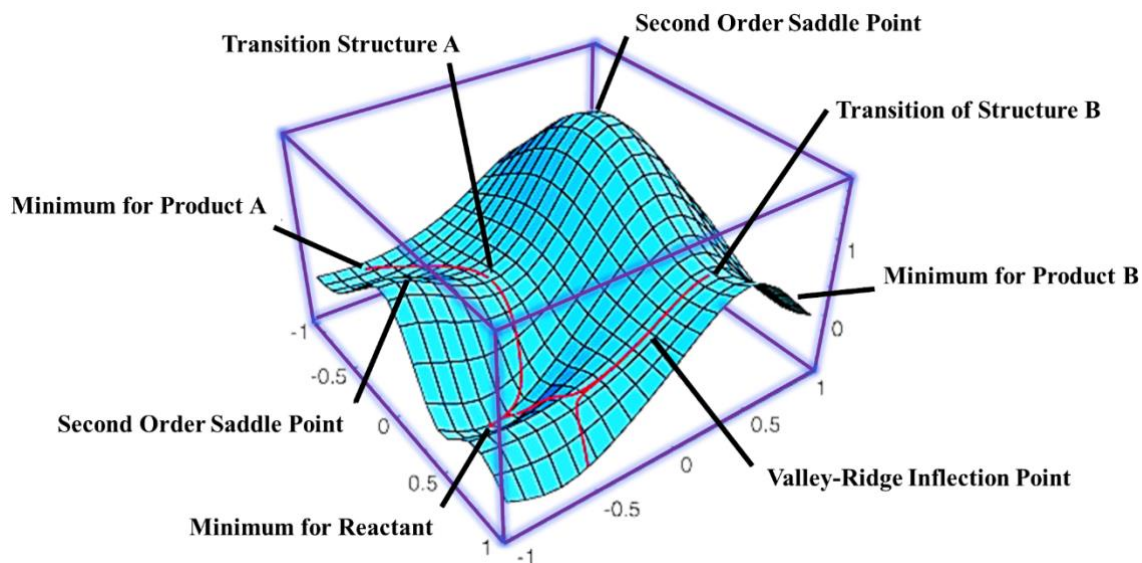
Where:  $\mathbf{H}_{\text{EN}}$  signifies the adjustments of terms-based activity to stabilise the nuclear positions ( $\mathbf{V}_{\text{NN}}$ ). Equation 3.2.5 represents  $\mathbf{E}_{\text{EN}}$ , deduced from the co-ordinates of the fixed nuclei and the unstable electrons.

$$(\mathbf{H}_{\text{el}} + \mathbf{V}_{\text{NN}})\Psi(\mathbf{r}_{\text{el}}) = \mathbf{E}_{\text{EN}}\Psi(\mathbf{r}_{\text{el}}) \quad (\text{Eq 3.2.6})$$

The electronic Schrödinger equation is used to describe molecular electronic motion. The accuracy of the approximation increases when applied to the electronic ground state. The fixed positions of interest concerning the equilibrated conformation may be evaluated once the equation has been solved. The potential energy surface and curve may also be established.

### ***3.2.3 Potential Energy Surface as an Application of Quantum Mechanics***

The potential energy surface (PES) is a valuable mathematical function that describes a molecule's energy, nuclei probability distribution and geometry by solving the Schrödinger equation (Figure 2). This concept emerged from the Born-Oppenheimer approximation, in which electron fluctuations are dependent on the positional states of the nuclei. The PES can then be taken as the potential of atom motion to collide into one another, within a molecule (Atkins and Friedman, 2011). The graphical illustration of PES (Figure 2) exhibits regions with high potential energy, signifying high-energy nuclear organisation or molecular conformations and low-energy regions that symbolize low nuclear energy conformations. The application of PES can be put to practical use in computational chemistry to assess the lowest state of energy and corresponding positional geometry of the molecule (Jensen, 2017).



**Figure 2:** Two-dimensional graphical representation of potential energy surface (Adapted from The California State University 2017).

### 3.3 The Principle of Molecular Mechanics

A major problem encountered in computational chemistry is the ability to efficiently comprehend the chemical characteristics of a compound, with regards to its stability, solubility and reactivity (Vanommeslaeghe and Guvench, 2014). To assess these features, the quantifiable dynamics must be observed and analysed using molecular-based models (Zimmermann *et al.*, 2017).

Molecular mechanics (MM) is simply defined as an approximation that simulates inter- and intra-molecular interactions within a molecule by algebraic, atomistic and empirical functions. The concept of MM, also referred to as the forcefield method, was intentionally developed to evaluate the energies and geometries of large biomolecular systems using a microscopic mechanical model. By doing so, the simulation costs were greatly reduced when dealing with large systems but require multiple molecular dynamic calculations (Vanommeslaeghe and Guvench, 2014). A molecule is considered by MM to be a collection of balls (the atoms) being held together by springs (designating the bonds). Within the structure model, the molecule's energy changes in response to its geometry due to the atoms repelling one another and bonds bending away from its natural angle or resisting the urge to be stretched. The application of MM integrates Newton's equation of motion for particles such as atoms, to describe a diverse array of molecular systems ranging from low molecular mass such as hydrocarbons to larger biomolecular complexes including nucleic acids and proteins (Vanommeslaeghe and Guvench, 2014).



In MM, the total energy of a compound is defined using simple algebraic terms, whilst neglecting the need to calculate the density of electrons and wave function, as observed in QM (Tsai, 2007). Various techniques are employed in RDD to determine potential drug leads prior to experimental testing and validation procedures. Molecular mechanic simulations also enable the development of atom models based on favourable energy projections (Poltev, 2017).

### 3.3.1 Potential Energy Function

In the force field method, atoms are designated as “building blocks” and electrons are not observed to be individual particles. This means that rather than resolving the Schrödinger equation, comprehensive information regarding the bonds should be provided instead. As mentioned previously, molecules are thought of as a “ball and spring” model with bonds of varying lengths and atoms of different sizes. It has been observed that different molecules may share structural similarities due the atomic make-up of the molecule. This concept was termed “atom type” and it relies on the type of bonding holding the atom in place and the atom number.

The force field/potential energy function of the molecular system may be distinguished by a set of force field equations that are based on Newton’s equation of motion. These equations can be used to calculate the energy of a system as well as the atomic types that create the overall molecule.

The total potential energy encompasses all potential inter- and intra- molecular elements including:

1. Bond stretching (occurring between atoms that directly interact with one another)

$$E_r = \sum K_r (r - r_0)^2 \quad (\text{Eq 3.3.1.1})$$

2. Angle bending (relevant atoms bound to the same central atom)

$$E_\theta = \sum K_\theta (\theta - \theta_0)^2 \quad (\text{Eq 3.3.1.2})$$

3. Bond torsion

$$E_\phi = \sum K_\phi [1 + \cos (\alpha_\phi - \Phi_0)] \quad (\text{Eq 3.3.1.3})$$

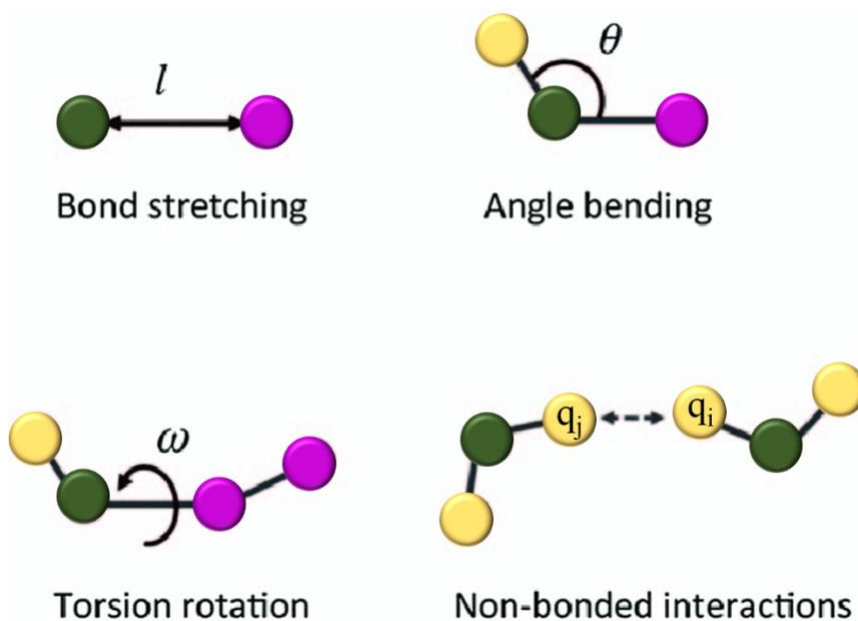
4. Non-bonded interaction (electrostatic and van der Waal forces)

$$E_{nb} = \left[ \sum \left( \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) \right] + \left[ \sum \left( \frac{q_i q_j}{D r_{ij}} \right) \right] \quad (\text{Eq 3.3.1.4})$$

Where the following represents:  $K_r$ ,  $K_\theta$ ,  $K_\phi$  are force constants for bond, angle, and dihedral angle, respectively. The equilibrium distance, angle and phase angle are signified by  $r_0$ ,  $\theta_0$ ,  $\Phi_0$ . Parameter  $r_{ij}$  is distance and  $A_{ij}$  and  $B_{ij}$  are van der Waal parameters. The molecular dielectric constant is represented by  $D$  and the charge points are  $q_i$  and  $q_j$ .

Typically, bonds are treated as springs and atoms as spheres in MM. The properties mentioned above are easier to mathematically elucidate when the atoms are treated as spheres with characteristics radii. The final PEF equation is therefore presented as:

$$E_{total} = E_r + E_\theta + E_\phi + E_{nb} \quad (\text{Eq 3.3.2})$$



**Figure 3:** Diagrammatic illustration of the total potential energy function of a molecule (Adapted from Saenz-Mendez *et al.*,2017).

Currently, there are several different force fields in use; however, they vary by the type of data used for fitting parameters, number of cross terms and function of each energy term. The two most common trends that are eminent when designing force fields:

1. Force field applied to large system such as protein or DNA, possess a relatively simplified functional form, void of cross terms and utilises the Lennard-Jones potential as van der Waals energy. These are known as diagonal or harmonic force fields.
2. Force fields applied on smaller to medium systems must uphold a high standard of accuracy. These types of force fields have several cross terms and an exponential-type potential for van der Waals energy. These are known as “Class II” force fields.

The most popular and wide spread used force fields are ENCAD (Vanommeslaeghe and Guvench, 2014), OPLS-AA (Sweere and Fraaije, 2017), CHARMM (Woodcock *et al.*, 2007; Galindo-Murillo, Roe and Cheatham, 2015), GROMOS (Schmid *et al.*, 2011; Reif, Winger and Oostenbrink, 2013) and AMBER (Case *et al.*, 2005, 2014). For purpose of this study, the harmonic AMBER force field was used for molecular system characterisation.

### **3.4 The Hybrid Quantum Mechanics/Molecular Mechanics**

The MM approach is extremely useful when investigating non-covalent interactions of large biomolecular systems, this method is however not adequately equipped to handle covalent systems as it does not allow effective reactivity. Quantum mechanics is an essential approach that primarily focuses on the outcome and mechanism of chemical interactions within small model systems (Vanommeslaeghe and Guvench, 2014). Together, these two approaches join forces to create an approach that combines the speed of MM and the accuracy of QM to allow the understanding of chemical interactions that take place within large biomolecular systems (Borbulevych, Martin and Westerhoff, 2018).

The hybrid QM/MM method was first introduced in 1976 by two scientists, Warshel and Levitt (Warshel and Levitt, 1976; Tiziana, Russo and Prejano, 2018). In this approach, small molecules such as residues of interest or atoms are exposed to the QM method to allow formation of chemical interactions while the remaining substrate regions and solvent are coherently subjected to the MM method (Cao and Ryde, 2018). The advantage of employing this approach over the conventional QM approach is that effect of the surrounding solvent and protein is also considered, in terms of their influence on the geometry and local electrostatic landscape (Lonsdale and Ward, 2018).

The QM/MM approach is often utilised to reveal the underlying mechanism of enzymatic activities, this can be beneficial in drug design, especially in relation to covalent compounds (Arodola and Soliman, 2017; Lonsdale and Ward, 2018). The hybrid approach is widely used and is very efficacious in simulation studies investigating the chemical reactivity, electronic energy

contributions, biochemical pathways and energy profiles of biomolecular systems (Lu *et al.*, 2016). Although the hybrid QM/MM approach is fast, efficient and highly cost-effective, advanced description by QM region has been adamantly circumvented due to the general lack of interpretation from reliable sources (Tuñón and Moliner, 2016). Additionally, the free energy calculations involved in this approach demand vast configurational sampling. Thus, with biomolecular system requiring longer molecular dynamic (MD) simulations, QM/MM method is not a suitable option. For this study, a longer MD simulation was needed, hence the all-atom MM covalent approach presented as the most fitting and feasible choice. All atom covalent simulations has posed quite a challenge in the scientific community due to the evident lack of literature supporting the appropriate implementation and execution of this technique. Therefore, we present a cost-efficient strategy for describing covalent interactions (S Khan *et al.*, 2018).

### **3.5 The Principle of Molecular Dynamic Simulations**

Molecular dynamics (MD) was first introduced by Rahman, Alder and Wainwright in the 1950s where it was implemented for the study of liquid dynamics (Alder and Wainwright, 1957,1959). Since then, the field of computational chemistry has substantially advanced allowing the method to progress from initially only being able to simulate hundreds to thousands of atoms, to now being capable of simulating entire systems involving macromolecules such as proteins and DNA.

There are two major types of simulation methods: classical MD and Monte Carlo (MC). A major advantage that reveals MD as the favoured method is the fact that it holds the ability to facilitate system dynamic characteristics such as time-dependent responses and rheological properties. Molecular dynamics is a treasured tool in molecular biology and biochemistry as it offers the opportunity to recognize and classify, from an atomistic perspective, the dynamic proceedings that may affect system properties.

Classical MD integrates Newton's equations of motion into its computational algorithm. At an atomistic level, MD simulations deliver high probability real-time mechanistic and conformational observations of chemical interactions, established through advance physical and mathematical algorithms. Implementation of MD simulations allows the exploration of particle interactions within a system, during a given time-period by generating a dynamic-based trajectory that is thereafter analysed. The main purpose of MD simulations is to incorporate Newton's equations to solve and explain the structural dynamics and energy contributions in a molecular network of a system. Initial particle conditions that must be adhered to are as follows:

1. Boundary conditions that need to be imposed
2. A competent force field for characterisation of interactive forces between atom, e.g. AMBER or CHARMM
3. Positions and velocities of individual particles

The conventional equation of motion may be solved as follows:

$$\mathbf{F}_i = \mathbf{m}_i \frac{d^2 \mathbf{r}_i(t)}{dt^2} \quad (\text{Eq 3.4})$$

Where:  $\mathbf{F}_i$  represents the interactive force acting on the particle,  $m$  is particle mass,  $t$  is time-evolution and  $\mathbf{r}_i$  is the position vector of the particle.

The MD process is divided into four continuous technical steps, which can be repeated numerous times to generate a trajectory. The steps involved are as follows:

1. The basic states of the biomolecular system are explained as
  - The co-ordinates of each atom
  - Bond properties between each atom
  - The accelerations of atoms
2. The potential energy of each atom is computed.
3. Energies obtained from step 2 are subsequently used to solve the equations of motion.
4. The system's new "state" needs to be retained and atom co-ordinates altered to step forward in the simulation. The process is then cycled back to step 1.

When the trajectory is fully generated, quantitative assessment of the system's evolution over the desired time-period may proceed.

### ***3.5.1 Covalent Molecular Dynamic Simulation***

Computational-based methods can describe the relationship between selective covalent inhibition and conformational alterations, thus offering a more reliable approach for covalent system exploration, opposed to standard simulation methods. Covalent molecular dynamic (CMD) simulations emerged as a result of a technical modification in the conventional MD method allowing researchers to overcome challenges arising from covalent bond parameterisation.

The first standard, reliable protocol for covalent simulation was proposed by Khan *et al.*, in 2018 (S Khan *et al.*, 2018). The protocol provided a valuable technical guideline rationalising covalent bond formation and corresponding system conformational changes. This strategy emphasised the point at which the two components, being the covalent drug and target protein, meet and interact. This technique has been successfully validated by several recent publications (S Khan *et al.*, 2018; Shama Khan *et al.*, 2018; Shunmugam and Soliman, 2018). The newly developed protocol includes suitable protein and ligand retrieval, system preparation and the final step of MD simulation utilising the all-atom MM approach.

Not all candidates are suitable when it comes to covalent system studies, therefore it is extremely important to maintain a specific selection criteria when choosing a ligand and its respective target protein. Selection of the protein and ligand depends on the presence of reactive moieties, also referred to as “warheads”. These moieties are described in accordance to their action. Functional groups that receive electrons are deemed electrophiles and those that donate are known as nucleophiles. The process involved in covalent bond formation is known as Michael addition reaction. This chemical reaction takes place between a Michael acceptor and donor. The residues involved in covalent bond formation includes glutamate, aspartate, histidine, arginine, threonine, tyrosine, lysine and serine but the most common of all, is cysteine. The unprotonated state of these residues are instrumental in the process of covalent inhibition.

### **3.5.2 Molecular Dynamics Post-Analysis**

Trajectories are created from the production run of the MD simulation. These trajectories are described as sequenced snapshots. They are characterised by velocity vectors and positional coordinates and document the time-evolution in a multidimensional space.

When selecting analytical software, there are three essential requirements:

1. Qualitative visualisation software that can produce high resolution images/snapshots and trajectory video clips.
2. The software should have swift processors that can handle large volumes of data.
3. A broad variety of analysis options should be available on each program.

The post-dynamic methods and calculations of interest are dependent on the nature of the MD study; while qualitative assessment is essential for system visualisation support.

For the purpose of this study, post-dynamic analysis of the trajectories are important for observation of:

- Conformational and energetic stability of the biomolecular system.
- Characteristics of the ligand's binding environment and the thermodynamic energy fluctuations along the clustered trajectory.
- Variability and dynamic geometric features of the biomolecular system.

### 3.5.2.1 System Stability

#### *Convergence*

The term “convergence” can be used to describe protein dynamics based on bond angle vibrations and bond type during the unravelling of a protein. This amalgamation toward equilibrium and the depiction of a conformational plateau and final energetics is crucial for a MD trajectory to be reproducible and accurate. It is at this plateau that the ligand-protein system is shown to exhibit energetically stable conformations (Amadei, Ceruso and Di Nola, 1999; Sawle and Ghosh, 2016).

#### *Root Mean Square Deviation (RMSD):*

The deviation experienced by a complex may be assessed by measuring the spatial difference arising between two static structures within the same trajectory. The RMSD of a trajectory is defined as follows:

$$\mathbf{RMSD} = \left( \frac{\sum_N (\mathbf{R}_i - \mathbf{R}_i^0)^2}{N} \right)^{\frac{1}{2}} \quad (\text{Eq 3.5})$$

Where: N is the total number of atoms in a complex,  $R_i$  is the vector position adopted by the  $C\alpha$  atom of particle i in the reference conformation, which is calculated following alignment of the structure to the original conformation (O) using the least square fitting. The average RMSD can be calculated by taking the average value over the number of frames in each trajectory and can be determined for a system's ligand, receptor and complex (Kufareva and Abagyan, 2012; Sargsyan, Grauffel and Lim, 2017).

#### *Radius of Gyration (RoG):*

The radius of gyration in a protein describes the root mean square distance of the atoms from their common centre of gravity of a given enzyme molecule. This analysis method allows the estimation of protein compactness along a given trajectory in a MD simulation. The equation below describes how RoG is assessed:

$$r^2 \mathbf{g} = \frac{\sum_{i=1}^n w_i (\mathbf{r}_i - \mathbf{r}^-)^2}{\sum_{i=1}^n w_i} \quad (\text{Eq 3.6})$$

Where:  $\mathbf{r}_i$  is the position of the  $i$ th atom and  $\mathbf{r}^-$  is the centre mass of atom  $i$ . The mean value is calculated by taking the RoG values over the number of frames in a given trajectory (Lobanov, Bogatyreva and Galzitskaya, 2008)

#### *3.5.2.2 Thermodynamic Energy Calculations (Binding Free Energy)*

Binding free energy (BFE) calculations is an imperative method utilised to observe the underlying binding mechanism between a ligand and protein, with enthalpic and entropic contributions (Schauperl *et al.*, 2017). The estimation of BFE initiates various approaches and algorithm developments including molecular docking calculations, linear interaction energy, integration of thermodynamics and free energy perturbation. Among the free energy calculations, the Molecular Mechanics/Poisson-Boltzmann Surface Area (MM/PBSA) and Molecular Mechanics/GB Surface Area (MM/GBSA) techniques manifest higher efficacy and accuracy abilities in estimating the binding free energies of biomolecular systems. The two energy terms are not reliant on a large training set to define different parameters. Both MM/PBSA and MM/GBSA use the combination of continuum solvent model and molecular model terms to estimate the absolute BFE, which is then averaged over the number of trajectory frames (Hayes and Archontis, 2012; Genheden and Ryde, 2015; Agoni, Ramharack and Soliman, 2018). The BFE ( $\Delta G$ ) of a biomolecular system is computed by the above-mentioned techniques and can be represented as:



$$(1) \Delta G_{\text{bind}} = \Delta G_{\text{complex}} - \Delta G_{\text{receptor}} - \Delta G_{\text{ligand}} \quad (\text{Eq 3.7.1})$$

$$(2) \Delta G_{\text{bind}} = E_{\text{gas}} + G_{\text{sol}} - T\Delta S \quad (\text{Eq 3.7.2})$$

$$(3) E_{\text{gas}} = E_{\text{int}} + E_{\text{vdw}} + E_{\text{ele}} \quad (\text{Eq 3.7.3})$$

$$(4) G_{\text{sol}} = G_{\text{GB}} + G_{\text{SA}} \quad (\text{Eq 3.7.4})$$

$$(5) G_{\text{SA}} = \gamma \text{SASA} \quad (\text{Eq 3.7.5})$$

Where:  $\Delta G_{\text{bind}}$  is summation of the gas-phase, gas-phase energy ( $E_{\text{gas}}$ ) and the solvation energy,  $G_{\text{sol}}$ , is less the entropy ( $T\Delta S$ ) term. The  $E_{\text{gas}}$  is the sum of internal energy,  $E_{\text{int}}$ , van der Waals (vdW) energy,  $E_{\text{vdw}}$  and electrostatic energy,  $E_{\text{ele}}$ . The total solvation energy is calculated by a summation of the total energy contributions of polar and non-polar states ( $G_{\text{GB}}$  and  $G_{\text{SA}}$ , respectively). The  $G_{\text{SA}}$  was determined from the solvent accessible surface area, generated by water probe radius of 1.4 Å. Energy determination of the polar state contributions was resolved by the  $G_{\text{GB}}$  equation. The total entropy of the solute is signified as ‘S’ and temperature as ‘T’. As per standard protocol, the solute and solvent dielectric constants are set to 1 and 80, respectively (Hou *et al.*, 2011; Zhang, Perez-Sanchez and C. Lightstone, 2017).

The MM/PBSA and MM/GBSA techniques suggest quantifiable analysis of ligand binding affinity to the protein and is therefore able to justify molecular docked structures with logical reasoning (Godschalk *et al.*, 2013; Sun *et al.*, 2018).

### 3.5.2.3 Conformational Features of System

#### Root Mean Fluctuation (RMSF):

The root mean fluctuation (RMSF) of a protein quantifies the C $\alpha$  atom fluctuations of each residue with regard to its average position along a given trajectory (Bornot, Etchebest and De Brevern, 2011). The RMSF analysis provides insight into the flexibility of various protein regions upon binding of a ligand. The following equation is used to calculate the standardised RMSF:

$$sRMSF_i = \frac{(RMSF_i - \text{RMSF})}{\sigma(\text{RMSF})} \quad (\text{Eq 3.8})$$

RMSF<sub>i</sub> represents the RMSF of the *i*th residue, from which the average RMSF is subtracted. This is then divided by the RMSF's standard deviation [ $\sigma(\text{RMSF})$ ] to yield the resultant standardised RMSF (sRMSF<sub>i</sub>).

#### 3.5.2.4 Dynamic Cross Correlation Matrix of a System (DCCM):

Dynamic cross correlation matrix (DCCM) is a method that's widely used to quantify fluctuations of residues in or out of phase during MD simulation. The co-efficient of cross correlation varies from anti-correlated motion (-1) to completely correlated motion (+1). The formula used to describe dynamic cross correlation is given below:

$$C_{ij} = \frac{\langle \Delta r_i \Delta r_j \rangle}{(\langle \Delta r_i^2 \rangle \langle \Delta r_j^2 \rangle)^{\frac{1}{2}}} \quad (\text{Eq 3.9})$$

The abbreviated terms denote the following: C<sub>ij</sub>: cross-correlation coefficient, fully correlated (-1) to anti-correlated (+1); *i*: *i*th residue; *j*: *j*th residue; Δ*r*<sub>*i*</sub>: displacement vectors correspond to *i*th residue and Δ*r*<sub>*j*</sub>: displacement vectors correspond to *j*th residue.

Dynamic cross correlation plots have become a very useful method for quantification of residue motion that result from ligand binding or in the event of protein mutation (Kasahara, Fukuda and Nakamura, 2014; Tiberti, Invernizzi and Papaleo, 2015).

### 3.6 Other Computer-Aided Drug Design Techniques Utilized in the Study

#### 3.6.1 Covalent Docking

Covalent docking has become a valuable tool in computational chemistry especially in structure-based drug design. The application of this docking form has garnered immense popularity due to the recent surge in covalent drug design and development (Shunmugam, Ramharack and Soliman, 2017). Conventional molecular docking methods calculates the binding mode of non-covalent inhibitors, but this technique is rather limited because it is not capable of forming a covalent bond. Moreover, steric resistance arising amongst the reactive residues and covalent functional group may prevent the recognition of appropriate docking conformations of the ligand. At present,

several covalent docking tools and platforms are available and extremely user friendly including DOCKTITE (Schmidt, 2014), DOCKoalent (London *et al.*, 2014), CovDock (Zhu *et al.*, 2014), Glide (Friesner *et al.*, 2006), GOLD (Verdonk *et al.*, 2003), AutoDock, (Cosconati *et al.*, 2010), CovalentDock (Ouyang *et al.*, 2013) and AutoDock Vina (Trott and Olson, 2010) (Shunmugam, Ramharack and Soliman, 2017). The covalent moiety must be loaded onto the software or online interface thereafter docking poses are generated in which the covalent bond is formed between the ligand and the target protein. Covalent docking has been successfully implemented in numerous studies, thus validating its important stature in CADD (S Khan *et al.*, 2018; Lonsdale and Ward, 2018).

### **3.6.2 Molecular Docking**

Molecular docking is a powerful tool employed in computational chemistry especially in drug design research. This technique utilises multiple methods to predict binding affinity and complex configuration. The most common example of molecular docking application is exhibited by ligand-receptor complexes and can also be used regarding protein-protein complexes as well as drug delivery complexes in association with aptamers or nanoparticles. Molecular docking is a two-step process: (1) Prediction and exploration of potential ligand binding conformations in addition to its orientation and position at the active site of the protein. (2) The evaluation of the binding affinity associated with each of the predicted binding poses (Meng *et al.*, 2011; Ferreira *et al.*, 2015).

In the last ten years, there has been a drastic increase in molecular docking studies. While they add value to the research domain regarding structural data of new compounds or novel biological targets, many inconsistencies have been encountered thus questioning the reliability of its use. This method has faced recurrent criticism related to the inappropriate selection of small molecules, inapt conformational poses and unsuitable binding sites (Ferreira *et al.*, 2015). In consequence to these concerns, all docked complexes in this study were further verified with MD simulations, followed by the demonstration of the ligand's stability at the active site.

### **3.6.3 Virtual Screening (VS)**

Virtual screening (VS) has a cardinal role in drug design and development research, this method is defined by the exploration of large-scale small molecule libraries in search of new compounds, based on the biological target of interest. These libraries generally contain millions of compounds and searching through each occupant will prove a difficult and tedious task. The VS method allows quick, effortless screening of these libraries, in order to filter through a smaller, more

feasible number of compounds that have the greatest chance as lead drugs. The objective of VS is achieved through recruitment of several filters, which are used to assist the identification of biologically active alternatives to current inhibitory compounds. This action is based largely on the “similar property principle”, that states that molecules with structural similarities tend to share similar properties.

Virtual screening is divided into three approaches: (1) Ligand-based VS (LBVS), (2) Structure-based VS (SBVS) and (3) Pharmacophore-based VS (PBVS). In the application of LBVS, compound libraries are generated using a well-known compound or its consequential binding interactions with a specific target. Contrastingly, SBVS identifies beneficial ligand binding affinities established at the targets active binding site. By doing so, it provides a new in-depth perspective into the behaviour of the active site region and ligand-protein interactions. This approach can positively identify selective compounds from immensely large compound libraries, which can then be subjected to further CADD studies. Even though both methods are innovative, their implementation has been associated with challenges, for instance, LBVS is known to generate substantially large compound libraries and identifying a possible lead drug may be a difficult task. Likewise, SBVS has inconsistencies such as solvation, receptor flexibility and entropic binding contributions, which poses a hurdle in the design of potential inhibitors against diseases and viruses. Recently, PBVS has gained a positive reputation accredited to its beneficial application in computational hit identification and lead optimisation. This form of VS is based on functional groups of known inhibitor compounds such as hydrophobic regions, aromatic rings, cations and hydrogen bond donors and acceptors. A criteria is drafted according to the pharmacophoric features of interest and thereafter, small molecule libraries are actively screened for possible hits. In this study, PBVS has been used, this approach to VS has been demonstrated by numerous studies, to be a much more reliable approach in comparison to LBVS and SBVS (Soliman, 2013; Appiah-Kubi and Soliman, 2017; Machaba *et al.*, 2017; Ramharack and Soliman, 2018).

## References

Agoni, C., Ramharack, P. and Soliman, M. E. S. (2018) ‘Co-inhibition as a strategic therapeutic approach to overcome rifampin resistance in tuberculosis therapy: Atomistic insights’, *Future Medicinal Chemistry*, 10(14), pp. 1665–1675. doi: 10.4155/fmc-2017-0197.

Amadei, A., Ceruso, M. A. and Di Nola, A. (1999) ‘On the convergence of the conformational coordinates basis set obtained by the Essential Dynamics analysis of proteins’ molecular dynamics simulations’, *Proteins: Structure, Function and Genetics*, 36(4), pp. 419–424. doi: 10.1002/(SICI)1097-0134(19990901)36:4<419::AID-PROT5>3.0.CO;2-U.

Appiah-Kubi, P. and Soliman, M. E. (2017) ‘Hybrid receptor-bound/MM-GBSA-per-residue energy-based pharmacophore modelling: enhanced approach for identification of selective LTA4H inhibitors as potential anti-inflammatory drugs.’, *Cell biochemistry and biophysics*, 75(1), pp. 35–48.

Arodola, O. A. and Soliman, M. E. S. (2017) ‘Quantum mechanics implementation in drug-design workflows: Does it really help?’, *Drug Design, Development and Therapy*, 11, p. 2551. doi: 10.2147/DDDT.S126344.

Atkins, P. W. and Friedman, R. S. (2011) *Molecular Quantum Mechanics*. Oxford university press.

Bahrami, M. *et al.* (2014) ‘The Schrödinger-Newton equation and its foundations’, *New Journal of Physics*, 16(11), p. 115007. doi: 10.1088/1367-2630/16/11/115007.

Baldi, A. (2010) ‘Computational approaches for drug design and discovery: An overview’, *Systematic Reviews in Pharmacy*, 1(1), p. 99.

Bohr, N. (1913) ‘On the constitution of atoms and molecules’, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 26(151), pp. 1–25.

Borbulevych, O., Martin, R. I. and Westerhoff, L. M. (2018) ‘High-throughput quantum-mechanics/molecular-mechanics (ONIOM) macromolecular crystallographic refinement with PHENIX/DivCon: the impact of mixed Hamiltonian methods on ligand and protein structure’, *Acta Crystallographica Section D: Structural Biology*, 74(11), pp. 1063–1077. doi: 10.1107/S2059798318012913.

Born, M. and Oppenheimer, R. (1927) ‘Zur quantentheorie der molekeln’, *Annalen der Physik*, 389(20), pp. 457–484.

Bornot, A., Etchebest, C. and De Brevern, A. G. (2011) 'Predicting protein flexibility through the prediction of local structures', *Proteins: Structure, Function and Bioinformatics*, 79(3), pp. 839–852. doi: 10.1002/prot.22922.

Cao, L. and Ryde, U. (2018) 'On the Difference Between Additive and Subtractive QM/MM Calculations', *Frontiers in chemistry*, 6, p. 89.

Case, D. a. *et al.* (2005) 'The Amber biomolecular simulation programs', *J. Computat. Chem*, 26(16), pp. 1668–1688. Available at: <http://onlinelibrary.wiley.com/doi/10.1002/jcc.20290/pdf>.

Case, D. A. *et al.* (2014) 'Amber 14'. doi: <http://www.ambermd.org>.

Cosconati, S. *et al.* (2010) 'Virtual Screening with AutoDock: Theory and Practice.', *Expert opinion on drug discovery*, 5(6), pp. 597–607. doi: 10.1517/17460441.2010.484460.

Doltsinis, N. L. (2006) 'Molecular Dynamics Beyond the Born-Oppenheimer Approximation : Mixed Quantum – Classical Approaches', *Computational Nanoscience: Do It Yourself!*, 31, pp. 389–409. doi: 10.3102/00028312038004915.

Ferreira, L. G. *et al.* (2015) 'Molecular docking and structure-based drug design strategies', *Molecules*, 20(7), pp. 13384–13421. doi: 10.3390/molecules200713384.

Friesner, R. A. *et al.* (2006) 'Extra precision glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes', *Journal of Medicinal Chemistry*, 49(21), pp. 6177–6196. doi: 10.1021/jm051256o.

Galindo-Murillo, R., Roe, D. R. and Cheatham, T. E. (2015) 'Convergence and reproducibility in molecular dynamics simulations of the DNA duplex d(GCACGAACGAACGAACGC)', *Biochimica et Biophysica Acta - General Subjects*, 1850(5), pp. 1041–1058. doi: 10.1016/j.bbagen.2014.09.007.

Genheden, S. and Ryde, U. (2015) 'The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities', *Expert Opinion on Drug Discovery*, 10(5), pp. 449–461. doi: 10.1517/17460441.2015.1032936.

Godschalk, F. *et al.* (2013) 'Comparison of MM/GBSA calculations based on explicit and implicit solvent simulations.', *Physical Chemistry Chemical Physics*, 15(20), pp. 7731–7739.

Goldbeck, G. (2012) *The Economic Impact of Molecular Modelling*, Goldbeck Consulting Ltd. doi:

10.5281/zenodo.44359.

Gorrec, F. (2014) 'Progress in macromolecular crystallography depends on further miniaturization of crystallization experiments', *Drug Discovery Today*, 19(10), pp. 1505–1507. doi: 10.1016/j.drudis.2014.07.002.

Hanson A., Hernandez JJ, Shadrick WR, Frick DN. Identification and Analysis of Inhibitors Targeting the Hepatitis C Virus NS3 Helicase. *Methods Enzymol.* 511, 463–483 (2012).

Hayes, J. M. and Archontis, G. (2012) 'MM-GB ( PB ) SA Calculations of Protein-Ligand Binding Free Energies', in *Molecular Dynamics – Studies of Synthetic and Biological Macromolecules*. InTech. doi: 10.5772/37107.

Hou, T. *et al.* (2011) 'Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations', *Journal of Chemical Information and Modeling*, 51(1), pp. 69–82. doi: 10.1021/ci100275a.

Jensen, F. (2017) *Introduction to computational chemistry*. John wiley & sons.

Kasahara, K., Fukuda, I. and Nakamura, H. (2014) 'A novel approach of dynamic cross correlation analysis on molecular dynamics simulations and its application to Ets1 dimer-DNA complex', *PLoS ONE*, 9(11), p. e112419. doi: 10.1371/journal.pone.0112419.

Katebi, A. R. *et al.* (2015) 'The Use of Experimental Structures to Model Protein Dynamics', *Methods in Molecular Biology*, 1215, pp. 213–236. doi: 10.1007/978-1-4939-1465-4.

Khan, S. *et al.* (2018) 'Covalent Simulations of Covalent/Irreversible Enzyme Inhibition in Drug Discovery – A Reliable Technical Protocol', *Future medicinal chemistry*, 10(19), pp. 2265–2275.

Khan, S. *et al.* (2018) 'Reversible versus irreversible inhibition modes of ERK2: A comparative analysis for ERK2 protein kinase in cancer therapy', *Future Medicinal Chemistry*, 10(9), pp. 1003–1015. doi: 10.4155/fmc-2017-0275.

Kore, P. P. *et al.* (2012) 'Computer-Aided Drug Design: An Innovative Tool for Modeling', *Journal of Medicinal Chemistry*, 2(1), pp. 139–148.

Kozlíková, B. *et al.* (2017) 'Visualization of Biomolecular Structures: State of the Art Revisited',

*Computer Graphics Forum*, 36(8), pp. 178–204. doi: 10.1111/cgf.13072.

Kufareva, I. and Abagyan, R. (2012) ‘Methods of protein structure comparison’, *Homology Modeling*, pp. 231–257. doi: 10.1007/978-1-61779-588-6\_10.

Lindahl, E. (2015) ‘Molecular Dynamics Simulations’, in *Molecular modeling of proteins: Second edition*. Springer Science & Business Media, pp. 3–26.

Lobanov, M. Y., Bogatyreva, N. S. and Galzitskaya, O. V. (2008) ‘Radius of gyration as an indicator of protein structure compactness’, *Molecular Biology*, 42(4), pp. 623–628. doi: 10.1134/S0026893308040195.

London, N. *et al.* (2014) ‘Covalent docking of large libraries for the discovery of chemical probes’, *Nature chemical biology*, 10(12), p. 1066. doi: 10.1038/nchembio.1666.

Lonsdale, R. and Ward, R. A. (2018) ‘Structure-based design of targeted covalent inhibitors’, *Chemical Society Reviews*, 47(11), pp. 3816–3830. doi: 10.1039/c7cs00220c.

Lu, X. *et al.* (2016) ‘QM/MM free energy simulations: recent progress and challenges’, *Molecular simulation*, 42(13), pp. 1056–1078.

Machaba, K. E. *et al.* (2017) ‘Tailored-pharmacophore model to enhance virtual screening and drug discovery: a case study on the identification of potential inhibitors against drug-resistant *Mycobacterium tuberculosis* (3R)-hydroxyacyl-ACP dehydratases’, *Future Medicinal Chemistry*, 9(10), pp. 1055–1071. doi: 10.4155/fmc-2017-0020.

Malhado, J. P., Bearpark, M. J. and Hynes, J. T. (2014) ‘Non-adiabatic dynamics close to conical intersections and the surface hopping perspective’, *Frontiers in Chemistry*, 2, p. 97. doi: 10.3389/fchem.2014.00097.

Meng, X.-Y. *et al.* (2011) ‘Molecular Docking: A Powerful Approach for Structure-Based Drug Discovery’, *Current Computer Aided-Drug Design*, 7(2), pp. 146–157. doi: 10.2174/157340911795677602.

Mohan, S. B. *et al.* (2012) ‘Computational Approaches for Drug Design and Discovery Process’, *Current Pharma Research*, 2(3), p. 600.



Ouyang, X. *et al.* (2013) ‘CovalentDock: Automated covalent docking with parameterized covalent linkage energy estimation and molecular geometry constraints’, *Journal of Computational Chemistry*, 34(4), pp. 326–336. doi: 10.1002/jcc.23136.

Poltev, V. (2017) ‘Molecular Mechanics: Principles, History, and Current Status’, in *Handbook of Computational Chemistry*. Springer, Cham, pp. 21–67. doi: 10.1007/978-3-319-27282-5\_9.

Ramharack, P. and Soliman, M. E. S. (2018) ‘Zika virus NS5 protein potential inhibitors: an enhanced in silico approach in drug discovery’, *Journal of Biomolecular Structure and Dynamics*, 36(5), pp. 1118–1133. doi: 10.1080/07391102.2017.1313175.

Reif, M. M., Winger, M. and Oostenbrink, C. (2013) ‘Testing of the GROMOS force-field parameter set 54A8: Structural properties of electrolyte solutions, lipid bilayers, and proteins’, *Journal of Chemical Theory and Computation*, 9(2), pp. 1247–1264. doi: 10.1021/ct300874c.

Saenz-Mendez, P. *et al.* (2017) ‘Computational chemistry and molecular modelling basics’, in *Computational Tools for Chemical Biology*, pp. 1–38. doi: 10.1039/9781788010139-00001.

Sargsyan, K., Grauffel, C. and Lim, C. (2017) ‘How Molecular Size Impacts RMSD Applications in Molecular Dynamics Simulations’, *Journal of Chemical Theory and Computation*, 13(4), pp. 1518–1524. doi: 10.1021/acs.jctc.7b00028.

Sawle, L. and Ghosh, K. (2016) ‘Convergence of Molecular Dynamics Simulation of Protein Native States: Feasibility vs Self-Consistency Dilemma’, *Journal of Chemical Theory and Computation*, 12(2), pp. 861–869. doi: 10.1021/acs.jctc.5b00999.

Schauperl, M. *et al.* (2017) ‘Binding Pose Flip Explained via Enthalpic and Entropic Contributions’, *Journal of Chemical Information and Modeling*, 57(2), pp. 345–354. doi: 10.1021/acs.jcim.6b00483.

Scherrer, A. *et al.* (2017) ‘On the mass of atoms in molecules: Beyond the born-oppenheimer approximation’, *Physical Review X*, 7(3), p. 031035. doi: 10.1103/PhysRevX.7.031035.

Schlick, T. (2013) *Molecular Modeling and Simulation: An Interdisciplinary Guide*. Springer Science & Business Media.

Schmid, N. *et al.* (2011) ‘Definition and testing of the GROMOS force-field versions 54A7 and 54B7’, *European Biophysics Journal*, 40(7), pp. 843–856. doi: 10.1007/s00249-011-0700-9.

Schmidt, B. (2014) 'DOCKTITE - A Highly Versatile Step-by-Step Workflow for Covalent Docking and Virtual Screening in MOE.', *Journal of Chemical Information and Modeling*, p. 141226132628000. doi: 10.1021/ci500681r.

Schrödinger, E. (1926) 'An undulatory theory of the mechanics of atoms and molecules.', *Physical review*, 28(6), p. 1049.

Shunmugam, L., Ramharack, P. and Soliman, M. E. S. (2017) 'Road Map for the Structure-Based Design of Selective Covalent HCV NS3/4A Protease Inhibitors', *Protein Journal*. Springer US, 36(5), pp. 397–406. doi: 10.1007/s10930-017-9736-8.

Shunmugam, L. and Soliman, M. E. S. (2018) 'Targeting HCV polymerase: a structural and dynamic perspective into the mechanism of selective covalent inhibition', *RSC Advances*, 8(73), pp. 42210–42222.

Sliwoski, G. *et al.* (2014) 'Computational Methods in Drug Discovery', *Pharmacological Reviews*, 66(1), pp. 334–395.

Soliman, M. E. . (2013) 'A Hybrid Structure/Pharmacophore-Based Virtual Screening Approach to Design Potential Leads: A Computer-Aided Design of South African HIV-1 Subtype C Protease Inhibitors', *Drug Development Research*, 74(5), pp. 283–295.

Stocks, M. (2013) 'The small molecule drug discovery process - from target selection to candidate selection', in *Introduction to Biological and Small Molecule Drug Research and Development: Theory and Case Studies*. Elsevier, pp. 81–126. doi: 10.1016/B978-0-12-397176-0.00003-0.

Sun, H. *et al.* (2018) 'Assessing the performance of MM/PBSA and MM/GBSA methods. 7. Entropy effects on the performance of end-point binding free energy calculation approaches', *Physical Chemistry Chemical Physics*, 20(21), pp. 14450–14460. doi: 10.1039/c7cp07623a.

Sutcliffe, B. T. and Woolley, R. G. (2012) 'On the quantum theory of molecules', *Journal of Chemical Physics*, 137(22), p. 22A544. doi: 10.1063/1.4755287.

Sweere, A. J. M. and Fraaije, J. G. E. M. (2017) 'Accuracy Test of the OPLS-AA Force Field for Calculating Free Energies of Mixing and Comparison with PAC-MAC', *Journal of Chemical Theory and Computation*, 13(5), pp. 1911–1923. doi: 10.1021/acs.jctc.6b01106.

Tatulian, S. A. (2018) 'From the Wave Equation to Biomolecular Structure and Dynamics.', *Trends in Biochemical Sciences*, 43(10), pp. 749–751. doi: 10.1016/j.tibs.2018.06.007.

Tiberti, M., Invernizzi, G. and Papaleo, E. (2015) '(Dis)similarity Index to Compare Correlated Motions in Molecular Simulations', *Journal of Chemical Theory and Computation*, 11(9), pp. 4404–4414. doi: 10.1021/acs.jctc.5b00512.

Tiziana, M., Russo, N. and Prejano, M. (2018) 'QM cluster or QM/MM in computational enzymology: the test case of LigW-decarboxylase.', *Frontiers in chemistry*, 6, p. 249.

Todd, A., Anderson, R. and Groundwater, P. W. (2009) 'Rational drug design - Identifying and characterising a target', *Pharmaceutical Journal*, 283(7559), pp. 19–20.

Trott, O. and Olson, A. J. (2010) 'AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading', *Journal of Computational Chemistry*, 31(2), pp. 455–61. doi: 10.1002/jcc.

Tsai, C. S. (2007) *Biomacromolecules: Introduction to Structure, Function and Informatics*, *Biomacromolecules: Introduction to Structure, Function and Informatics*. John Wiley & Sons. doi: 10.1002/0470080124.

Tuñón, I. and Moliner, V. (2016) *Simulating Enzyme Reactivity: Computational Methods in Enzyme Catalysis*. Royal Society of Chemistry.

Vanommeslaeghe, K. and Guvench, O. (2014) 'Molecular Mechanics', *Current pharmaceutical design*, 20(20), pp. 3281–3292.

Verdonk, M. L. *et al.* (2003) 'Improved protein-ligand docking using GOLD', *Proteins: Structure, Function and Genetics*, 52(4), pp. 609–623. doi: 10.1002/prot.10465.

Wang, Z. (2016) *On the models of atom*.

Warshel, A. and Levitt, M. (1976) 'Cite Warshel, Arieh, and Michael Levitt. "Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme', *Journal of Molecular Biology*, 103(2), pp. 227–249.

Woodcock, H. L. *et al.* (2007) 'Interfacing Q-Chem and CHARMM to perform QM/MM reaction path

calculations’, *Journal of Computational Chemistry*, 28(9), pp. 1485–1502. doi: 10.1002/jcc.20587.

Zhang, X., Perez-Sanchez, H. and C. Lightstone, F. (2017) ‘A Comprehensive Docking and MM/GBSA Rescoring Study of Ligand Recognition upon Binding Antithrombin’, *Current Topics in Medicinal Chemistry*, 17(14), pp. 1631–1639. doi: 10.2174/1568026616666161117112604.

Zhu, K. *et al.* (2014) ‘Docking covalent inhibitors: A parameter free approach to pose prediction and scoring’, *Journal of Chemical Information and Modeling*, 54(7), pp. 1932–1940. doi: 10.1021/ci500118s.

Zimmermann, M. T. *et al.* (2017) ‘Molecular modeling and molecular dynamic simulation of the effects of variants in the TGFBR2 kinase domain as a paradigm for interpretation of variants obtained by next generation sequencing’, *PLoS ONE*, 12(2), p. e0170822.

## CHAPTER 4

### **Road Map for the Structure-Based Design of Selective Covalent HCV NS3/4A Viral Protease Inhibitors**

Letitia Shunmugama, Pritika Ramharacka, Mahmoud E. S. Soliman<sup>a,b\*</sup>

<sup>a</sup>Molecular Modeling and Drug Design Research Group, School of Health Sciences, University of KwaZulu-Natal, Westville Campus, Durban 4001, South Africa

<sup>b</sup>College of Pharmacy and Pharmaceutical Sciences, Florida Agricultural and Mechanical University, FAMU, Tallahassee, Florida 32307, USA.

\*Corresponding Author: Mahmoud E.S. Soliman

School of Health Sciences, University of KwaZulu-Natal, Westville Campus, Durban 4001, South Africa

Email: [soliman@ukzn.ac.za](mailto:soliman@ukzn.ac.za)

Telephone: +27 (0) 31 260 8048, Fax: +27 (0) 31 260 7872

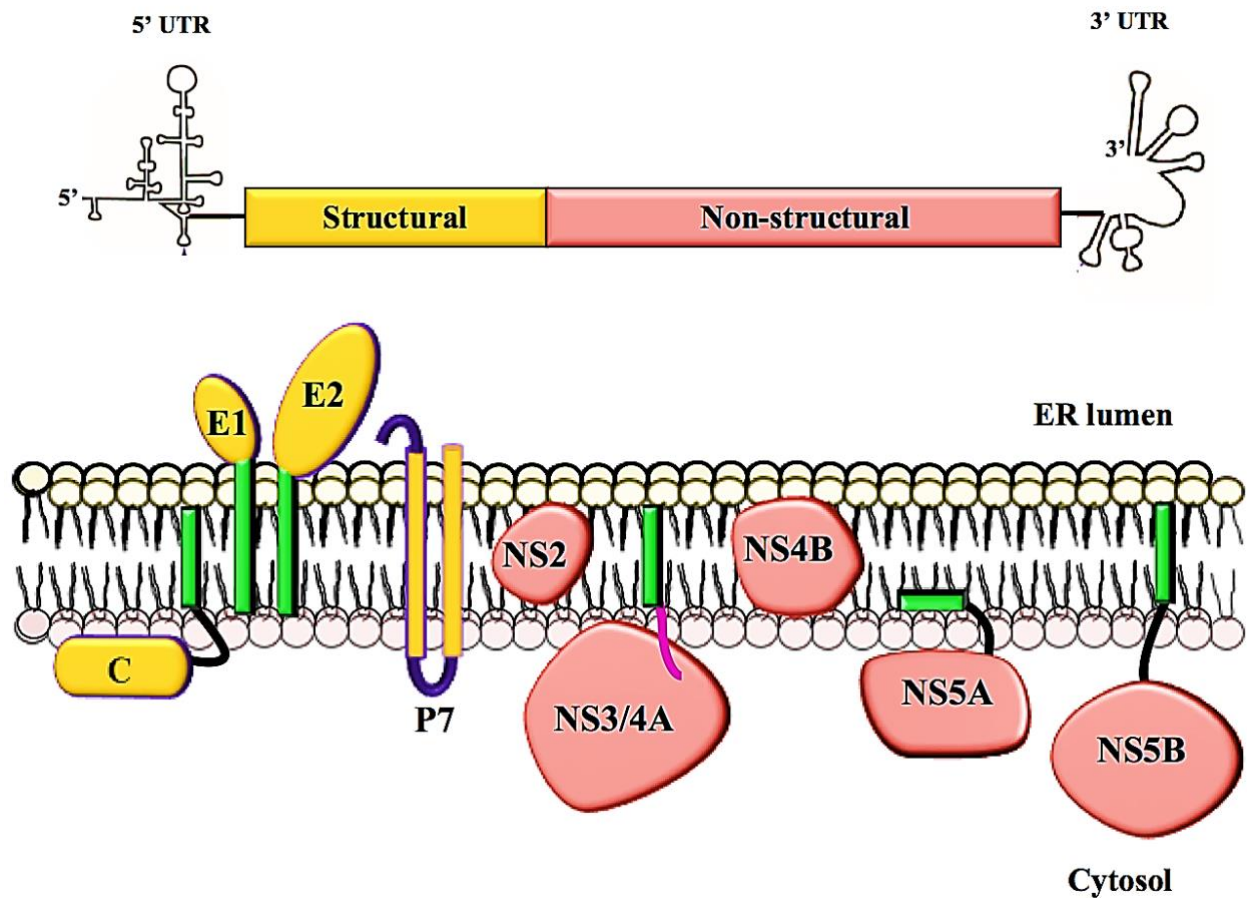
## **Abstract**

Over the last two decades, covalent inhibitors have gained much popularity and is living up to its reputation as a powerful tool in drug discovery. Covalent inhibitors possess many significant advantages including increased biochemical efficiency, prolonged duration and the ability to target shallow, solvent exposed substrate-binding domains. However, rapidly mounting concerns over the potential toxicity, highly reactive nature and general lack of selectivity have negatively impacted covalent inhibitor development. Recently, a great deal of emphasis by the pharmaceutical industry has been placed toward the development of novel approaches to alleviate the major challenges experienced through covalent inhibition. This has unexpectedly led to the emergence of “selective” covalent inhibitors. The purpose of this review is not only to provide an overview from literature but to introduce a technical guidance as to how to initiate a systematic “road map” for the design of selective covalent inhibitors which we believe may assist in the design and development of optimized potential selective covalent HCV NS3/4A viral protease inhibitors.

**Keywords:** Hepatitis C virus; NS3/4A viral protease; covalent inhibitors; structure-based drug design; selectivity

## 1 Introduction

Hepatitis C virus (HCV) has been deemed a global health concern as infection rates has escalated to between 170 and 185 million people worldwide <sup>1</sup>. The progression of chronic HCV has been shown to facilitate the development of viral hepatitis, hepatic cirrhosis and/or hepatocellular carcinoma, ultimately resulting in severe liver damage <sup>1-3</sup>. Belonging to the *Hepacivirus* genus and a member of the *Flaviviridae* family, HCV is a hepatotropic, positive-strand RNA virus which encodes a polyprotein that undergoes proteolytic cleavage to generate 4 structural proteins (C, E1, E2 and P7) and 6 non-structural proteins (NS2, NS3, NS4A, NS4B, NS5A and NS5B) <sup>4</sup>.



**Figure 1:** The schematic representation of HCV genomic polyprotein organisation. The 3000 kDa 5'-3' untranslated region (UTR) is comprised of 4 conserved domains which translate into 4 structural proteins: core (C), envelope proteins 1 and 2 (E1/E2) and hydrophilic ion channel protein 7 (P7). The NS3/4A protease mediates the cleavage of the HCV polyprotein resulting

in the generation of non-structural proteins: NS2, NS4A co-factor, interferon resisting NS5A protein and NS5B RNA dependent RNA polymerase <sup>4</sup>.

During the early stages of NS3/4A viral protease inhibitor research, high throughput screening/virtual screening technique was applied to search through vast numbers of small molecule libraries with the goal of generating potential lead compounds, thus speeding up the rational drug design process. In spite of this effort, no promising molecules were identified. The failed attempt at screening was followed by a strategic SBDD approach. In doing so, the approach dynamically prompted the discovery of non-covalent inhibitors. It wasn't until the 21<sup>st</sup> century, whereby drug discovery steered researchers on a path toward the discovery of covalent inhibitors. Ever since, the design and development of covalent inhibitors as potential drug candidates has gained great interest in the pharmaceutical industry <sup>5,6</sup>.

Covalent inhibition serves as a useful method in target identification and drug discovery <sup>7</sup>. Small organic molecules are utilised as covalent inhibitors which contain a moiety that is specifically designed to bind to an amino acid side chain of the target protein <sup>8</sup>. This occurs by way of conventional reversible or irreversible interactions whilst undergoing a bond-forming modification that intensifies the binding affinity between the covalent inhibitor and target protein <sup>9</sup>. Covalent irreversible modifications of proteins form non-labile persistent bonds with therapeutic benefits in comparison to traditional non-covalent reversible binding <sup>10</sup>. Many potent inhibitors have been designed and discovered to form covalent adducts with target residues found in the catalytic binding site of proteases. Research into HCV NS3/4A viral protease inhibitors have identified three distinct examples of covalent inhibition: telaprevir, boceprevir and narlaprevir. However, a high level of similarity is apparent amongst HCV and related viral protease active sites, thus rendering covalent inhibitors less than favourable with respect to selectivity <sup>11</sup>. There is an abundance of information on covalent HCV inhibitors however there are limited molecular approaches that can be used to assist the design of selective covalent inhibitors in literature. In this review, we provide an exclusive compilation of published works over the last 15 years which may provide technical guidance to assist in the deliberate design of future selective covalent inhibitors.

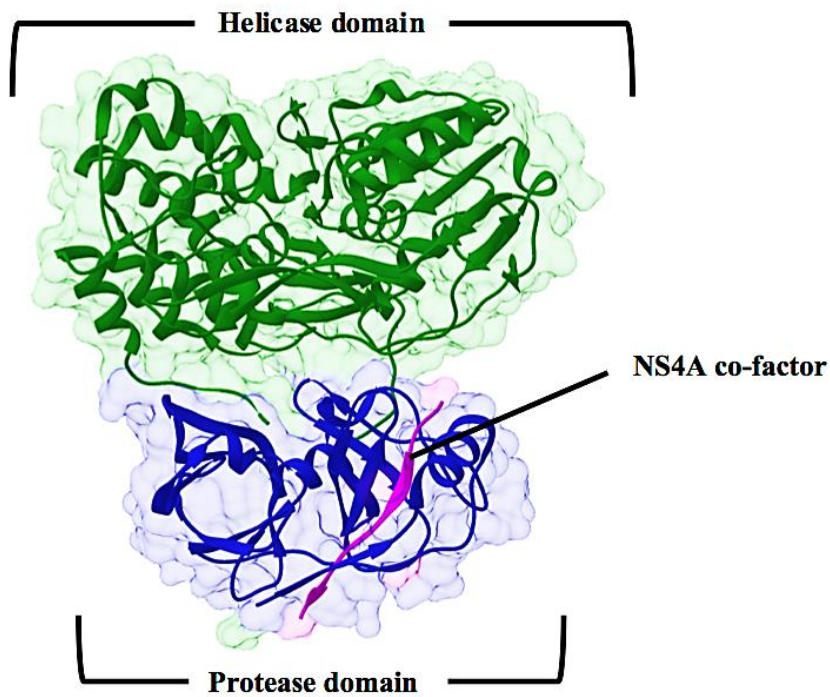
## **2 HCV protease as a critical target in drug discovery**

The catalytic protease domain of NS3/4A protease consists of residue 1-180 residue and represents the N-terminal of the 631 residue protein <sup>12</sup>. The ATP-dependent RNA helicase is encoded by the larger C-terminal of NS3 (residue 181-631) <sup>2,4,13</sup>. Initially, the catalytic protease domain of NS3 is inactive and requires co-factor NS4A, a 54 residue protein with a hydrophilic C-terminus and a hydrophobic N-terminus <sup>13,14</sup>. Upon the binding of NS4A, the NS3 domain undergoes rearrangement to facilitate the enzymatically proficient alignment of the catalytic triad (H57, D81, and S139)<sup>13</sup>. The NS4A co-factor is ultimately essential for the initiation of NS3 protease activity, allowing its localisation and

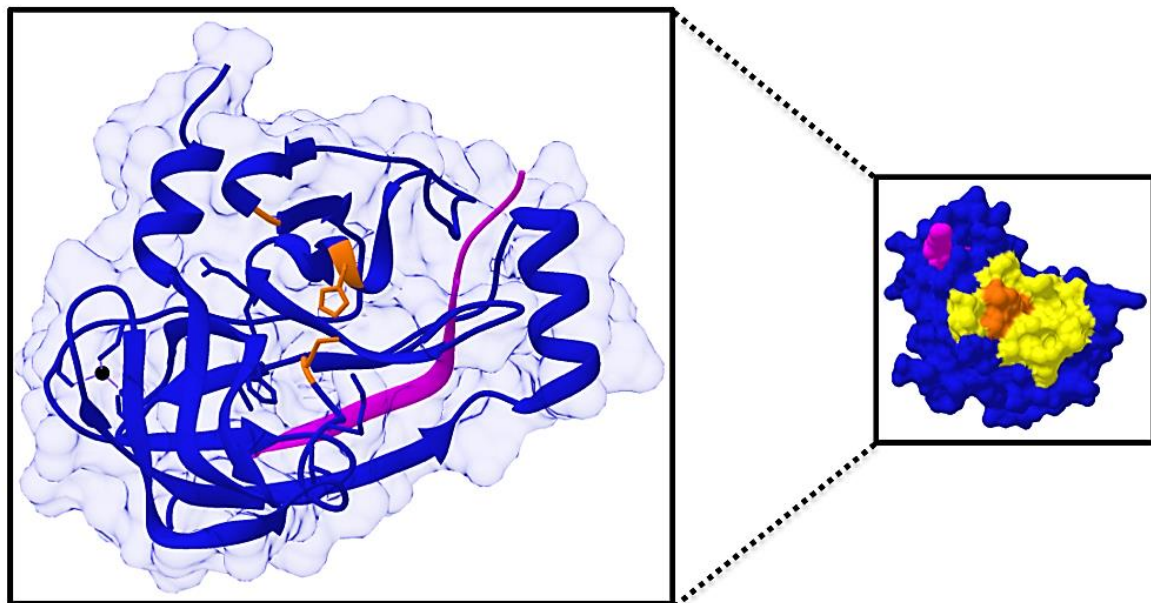


stabilisation at the endoplasmic reticulum (ER) membrane as well as cleavage-dependent activation of HCV biogenesis (Figure 2) 15.

**A: HCV NS3/4A bifunctional complex**



**B: Protease domain of HCV NS3/4A complex**



**Figure 2:** (A) The crystal structure of NS3/4A serine protease holoenzyme. (PDB ID: 4B6E). (B) Representation of the protease domain of NS3 in complex with NS4A (PDB ID: 4U01). The catalytic triad residues of the active site are positioned in the cleft which connects the

helicase and protease domains. The catalytic binding site is shallow as indicated in the insert. Zinc has been identified as an important structural feature however its exact function is unknown. Protease domain: blue; helicase domain: green; NS4A co-factor: magenta; catalytic triad: orange; active binding site: yellow; zinc: black

In the last 2 decades, research efforts have focused on virus-encoded proteases as potential targets for antiviral drug therapy. The viral proteases is an absolute necessity in the life cycle of HCV either by effecting the cleavage of precursor viral proteins to yield functional products or by mediating the processing of structural proteins required for assembly and morphogenesis of viral particles. Thus the functional importance of NS3/4A makes it an intensively studied viral protein that represents an attractive antiviral drug target. Initially the preclinical efficiency of protease inhibition was demonstrated in HIV-infected individuals. This finding nurtured hope that other viral proteases, such as HCV NS3/4A, could be an ideal target for a structure-based drug design approach (SBDD).

### **3 Non-covalent vs covalent inhibition of HCV NS3 protease**

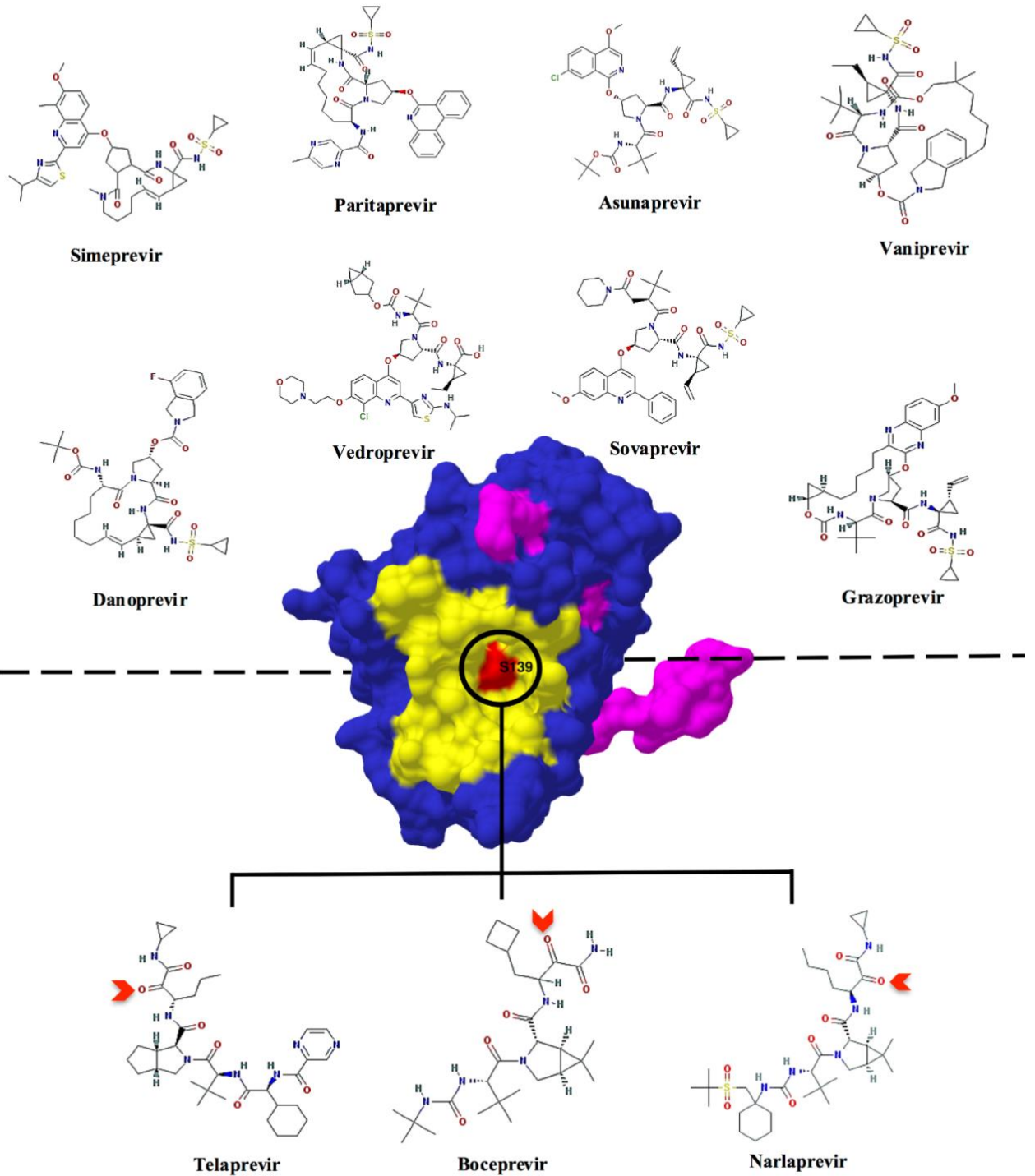
Although immense information was available within literature on the subject of HCV proteases there was still a staggering lack in both development and availability of effective HCV protease inhibitors. It wasn't until 1996, where a revolutionary breakthrough by Kim *et al* (1996) led to the publication of the HCV NS3/4A viral protease crystal structure <sup>16</sup>. This ignited intense research efforts among pharmaceutical companies toward the identification of inhibitors for therapeutic interventions <sup>17</sup>. Another massive influence that greatly impacted HCV drug discovery was the discovery of the protease's vulnerability to enzyme inhibition which was identified by means of peptide product release from the N-terminal following substrate cleavage <sup>18</sup>. These discoveries led to many years of exploitation in the direction of successful HCV protease inhibition through SBDD. Numerous inhibitors have been designed to non-covalently and covalently inhibit the HCV viral protease. Most often in the case of non-covalent protease inhibitors, they non-selectively dock into the catalytic/substrate binding pocket. Contrastingly covalent protease inhibitors specifically target a certain residue at the non-catalytic or catalytic domain of the protein of interest.

#### **3.1 Non-covalent HCV protease inhibition**

Once the crystal structure of HCV protease became public knowledge, relentless efforts by pharmaceutical industries were placed toward the development of several potential antiviral candidates including ciluprevir (BILN 2061). Ciluprevir (Boehringer Ingelheim, Germany) was the first product-based drug candidate to proceed to clinical drug trials in humans. The drug demonstrated a significant reduction of the HCV viral load thus establishing irrefutable evidence of HCV NS3/4A protease as good target for drug development <sup>19</sup>. Unfortunately, ciluprevir was discontinued due to heightened occurrence

of cardiac toxicity exhibited at high doses within animal models. Since ciluprevir was effectuous against HCV, it created a fundamental understanding about the interactions occurring between protease inhibitors and the HCV NS3/4A viral protease. In the years to follow, ciluprevir's chemical scaffold served as a template and was utilised in the development of at least 2 non-covalent HCV protease inhibitors, danoprevir and simeprevir <sup>20</sup> (Figure 3).

## Non-covalent Inhibitors



## Covalent Inhibitors

**Figure 3:** Examples of non-covalent and covalent inhibitors including their HCV NS3/4A protein target (PDB: 2OC8). All non-covalent and covalent chemical structures were obtained using

PubChem. Non-covalent inhibitors generally dock to the shallow binding pocket of NS3/4A protease and held in place by intermolecular forces. Contrastingly, covalent inhibitors specifically target and bind to the S139 residue located at the catalytic site of the HCV protease to initiate inhibition. The red arrows represent the inhibitor atom that undergoes covalent modification to facilitate the formation of a covalent bond between the inhibitor and S139. NS3 protease domain: blue; NS4A co-factor: magenta; active binding site: yellow; S139: red.

Apart from conventional non-covalent inhibitors, an additional strategy was adopted and implemented in SBDD of HCV inhibitors. The newly found approach involved the deliberate integration of substrate-based inhibitors with an electrophilic war-head, for instance an alpha-ketoamide functional group. The electrophilic war-head then acts as an active serine trap thus ultimately creating a “covalent” adduct or linkage between the inhibitor and target protein. This pioneering approach has led researchers on a promising path toward the discovery of target specific covalent HCV inhibitors.

### **3.2 Covalent HCV protease inhibition**

Covalent inhibitors of HCV are characterized by the formation of an adduct, which links the hydroxyl group of the protease’s catalytic serine residue to the electrophilic group of the inhibitor. The resulting covalent complex may be categorised as either “reversible” or “irreversible”<sup>20</sup>. Reversible covalent inhibitors covalently bind to their target protein, consequently disassociating from it at a rate greater than the turnover rate of protein synthesis<sup>21</sup>. Conversely, irreversible covalent inhibitors form adducts with their target protein which may not separate from the protein for the duration of its lifespan or do so with a kinetic half-life portentously longer than the rate of protein re-synthesis<sup>21</sup>.

In the scientific community, covalent inhibitors have been subjected to wide-scale dispute due to the discrepancies arising from the plethora of advantages compared to expanding list of disadvantages. There are numerous advantage of covalent inhibitors but the following top the list: increased biochemical efficacy initiated by non-equilibrium binding restricts the competition that may arise between elevated ligand or substrate concentrations; prolonged duration of drug action thereby requiring less frequent doses; increased potency; ability to bind to shallow substrate binding sites and to avoid mechanisms of<sup>6,19,21,22</sup>. The cons of covalent inhibitors range from idiosyncratic toxicity caused by target protein haptensisation leading to immune system initiating host antibody response, reactivity that offsets and prohibits any selectivity that causes toxic injury

ultimately resulting in the initiation of host cell damage responses and lastly immunogenicity caused by the formation of protein inhibitor adducts <sup>23,24</sup>.

In HCV drug development, the most detrimental adverse effects of covalent inhibitors are increased toxicity levels and off-target reactivity. However, recent studies have identified 3 distinct novel HCV protease inhibitors: telaprevir, boceprevir and narlaprevir that have re-centred the focus on covalent inhibitors <sup>6</sup>. All 3 covalent HCV protease inhibitors target the catalytic binding site, more specifically serine at position 139 (S139) (Figure 3). Following the binding of the covalent inhibitors to the HCV viral protease, the alpha-ketoamide forms a reversible covalent linkage with S139. Unfortunately, this particular serine residue is conserved across all proteases and is therefore susceptible to complications emanating from the lack of selectivity (Figure 4) <sup>11</sup>.

<b>HCV</b>	ISYLGKGS <b>S</b> GGPLLCPAGHAVGLFRAAV <b>C</b> T
<b>DV</b>	LDFSPGT <b>S</b> GSPIIDRKGKVVGLYGNGVVT
<b>WNV</b>	LDYPTGT <b>S</b> GSPIVDKNGDVIGLYGNGVIM
<b>ZIKA</b>	LDYPAGT <b>S</b> GSPILDKCGRVIGLYGNGVVI *

**Figure 4:** Alignment of NS3 viral protease sequence from 4 members of the *Flavivirus* family. S139 is conserved within most proteases however C159 is only present in the HCV NS3 protease sequence. DV: Dengue virus; WNV: West Nile virus; S: serine 139 (red); C: cysteine 159 (blue); \* indicates identical residues.

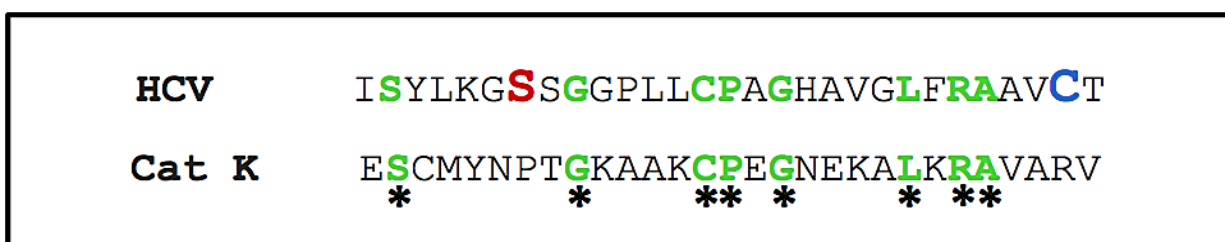
#### 4 Exploring the concept of “selective” covalent inhibition

The design of selective covalent inhibitors is appealing concept but in practice, it is considerably harder to achieve. This may primarily be caused by the difficulty arising in maintaining the correct balance between selectivity and reactivity <sup>23</sup>.

Hagel and co-investigators (2011) were the first to report on a protease inhibitor forming a covalent linkage with a targeted cysteine residue located at the non-catalytic site, outside the

designated active site <sup>11,25</sup>. The targeted cysteine residue was identified as nucleophilic C159 and it is conserved across all 919 HCV NS3 sequences to date. Even though structural similarities are observed between HCV NS3/4A and the host proteases, C159 is structurally unique to only the HCV NS3/4A viral protease (Figure 4 and 5). Thus it was determined that C159 was an ideal target for obtaining selectivity between HCV and host proteases <sup>11</sup>.

The researchers in this study were able to experimentally design and identify compound 3 (IC<sub>50</sub> = 2nM) as a potent selective covalent HCV viral protease inhibitor. Compound 3 contains an electrophilic acrylamide via a D-alanine linker which aids in increased conformational restraint of the molecule. To validate the findings of the selective covalency of compound 3, Hagel *et al* (2011) tested the compound against a NS3 viral protease harbouring a serine substitution at position 159 (C159S) <sup>11</sup>. The results revealed compound 3 was unable to bind covalently to C159S thereby confirming its unequivocal selectivity for C159 of the HCV viral protease <sup>11</sup>.



**Figure 5:** Sequence alignment comparing HCV viral protease with Homo sapien derived cathepsin K protease. Cathepsin K is 1 of 500 proteases found in Homo sapiens and is functional in bone resorption. As depicted, there are some structural similarities (green) however cathepsin K lacks S139 (red) and C159 (blue). \* indicates identical residues.

### 5 *In silico* approach for the design of selective covalent HCV inhibitors

In this section, we provide technical guidance as to how to initiate a systematic road map for the design of selective covalent inhibitors. These are typical procedures however some of the steps require further modification in accordance to the protein of interest, inhibitor under investigation and the resources available.

#### 5.1 Target Identification

The first step is target identification. There are many bioinformatic resources available that can be utilised for this step. Characteristically, the acquisition of the target protein's crystal structure,



in this case, HCV NS3/4A viral protease complex can be obtained from RCSB protein data bank 26. If the crystal structure of the target protein is not available, a proposed structure can be modelled through bioinformatic and homology modeling software (Table 1). Secondly, it is imperative to identify the substrate-binding domain including active site residues of the target protein. This can be extrapolated from literature or identified using programs specialised in substrate-binding domain identification (Table 1).

## 5.2 Ligand recognition

The second step of designing a selective covalent inhibitor is to identify the ligand or inhibitor of interest. There are several identification approaches available however in our research laboratory, we rely on 2 main ligand identification approaches. The first approach encompasses the direct modification of published HCV protease inhibitors, through the addition of a covalent war-head by means of molecular modeling. The apposite position of the covalent war-head on the inhibitor molecule is dictated by prior literature pertaining to the binding mode with the target protein's non-catalytic C159 residue. Alternatively, the second approach involves the use of high throughput screening (HTS) also known as virtual screening, to identify a set of non-covalent HCV viral protease inhibitor templates. The templates will subsequently be converted to potential covalent inhibitors through molecular modeling, followed by the structural addition of a covalent war-head 26,27.

## 5.3 Covalent Docking

Covalent docking is considered a critical topic in literature and forms a challenging task. In our previous reports we have highlighted the merits and pitfalls associated with covalent docking 6. To date, multiple different computational tools can be implemented in covalent docking for the sole purpose of potential drug discovery. The most common approach is the "link atom" tactic. In this approach, a "link atom" is designated in both the inhibitor and target protein. This compels the link atom of the inhibitor to occupy the same spatial arrangement as the target protein's link atom in order to simulate covalent binding. Molecular docking software program such as GOLD implements this *in silico* approach 28. Another widely utilised molecular docking program is AutoDock Tools. This program applies two methods during covalent docking: (1) a "grid-based" method and (2) flexible side chain modification. In the grid-based method, an attachment atom of the target protein and grid-based energy are the focal aims of a Gaussian based function for proclivity of the inhibitor's covalently bound pose. With regard to the flexible side chain method, the covalently bound inhibitor and protein attachment are both perceived as one flexible side chain and experimented as part of the receptor 29. The shortcomings of these covalent docking

applications are the need for manual preparation of the inhibitor/ligand and target protein as well as the designation of the reactive atoms and type of reaction <sup>30</sup>. The 2 mentioned programs, Autodock Tools and GOLD may adequately function in covalent docking but have major constraints in their accuracy, functionality and overall usefulness. Additionally, since manual preparation is a prerequisite, the lack of automation in the experimental system can hinder the use of covalent docking progression. A molecular docking software program, CovalentDock, has addressed these issues by automating the preparation of ligand files <sup>6,30</sup>.

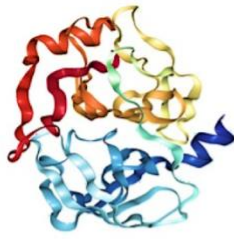
The basic mechanism involved in covalent interactions is the initial non-covalent binding of the inhibitor to the target protein. Non-covalent docking of the inhibitor can be achieved through virtual screening software, AutoDock Vina, whereby the inhibitor is docked into the target protein's specified non-catalytic binding site <sup>31</sup>. UCSF Chimera can then be utilised to acquire the values of the grid box parameters <sup>32</sup>. The resultant compound assumes a pose that provides beneficial optimum reaction for selective docking. The second step of the process is covalent docking by way of CovalentDock <sup>30</sup>. Once the binding affinities are established for the covalently docked ligand, the conformation with the highest binding affinity (lowest binding energy) should be selected for molecular dynamic (MD) simulations <sup>33</sup>. Other than CovalentDock, there are several additional molecular docking software programs available to assist with covalent docking (Table 1).

#### **5.4 MD: Quantum Mechanic/Molecular Mechanics Simulation**

*In silico* MD simulations are carried out to further the understanding of protein-ligand complex properties in terms of its molecular structure, intermolecular interactions and electronic rearrangement <sup>34,35</sup>. Quantum Mechanic/Molecular Mechanic (QM/MM) implementation is a newly emerging topic in drug discovery. Recently, in 2017, we have produced a report which has focused on and highlighted advantages associated with the application of QM/MM in drug design <sup>36</sup>. The QM/MM simulation is a hybrid technique that is widely held for the study of biomolecular system modeling and chemical interactions in proteins and within solutions. This approach combines the accuracy of quantum mechanics and the speed of molecular mechanics <sup>35,37</sup>. In QM/MM, the system's region in which harbours the chemical process is treated at an appropriate level of theoretical quantum chemistry whilst the remaining system is described by a molecular mechanical force field such as Amber <sup>35</sup>. Several other force fields may be applied, depending on the investigator's preference (Table 1).

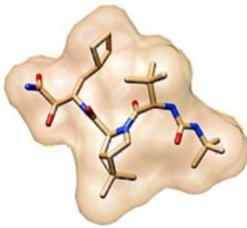
#### **5.5 Post Dynamics Analysis**

Post-dynamic analysis of the trajectories obtained from MD simulation is vital in defining the relative binding free energies, dynamic conformational attributes and configurations, and to determine the state of thermodynamic equilibrium of biomolecular systems <sup>38</sup>.



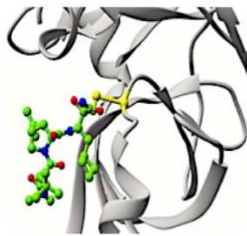
### Target protein identification

- Select a target protein
- Create 3D homology model if not available
- Identify substrate binding domains



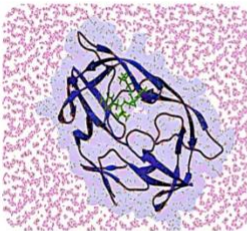
### Virtual screening- Ligand

- Structure/ligand based generation via compound library
- Physiochemical and biochemical filters (Lipinski's rule of five)



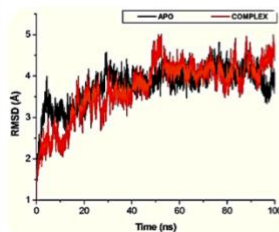
### Covalent Docking

- Dock inhibitor/s to target protein
- Define the residue of targeted importance
- Rank according to binding affinities
- Select highest binding affinity



### Molecular Dynamic Simulations

- Quantum Mechanics/Molecular Mechanics method



### Post Dynamic Analysis

**Figure 6:** Summarised workflow adopted *in silico* by selective covalent inhibitor drug design.

**Table 1:** *In silico* software available for SBDD of selective covalent HCV viral protease inhibitors.

<b><i>In silico</i> method</b>	<b>Software available</b>
<b>Target identification</b>	RCSB Protein DataBank <sup>39</sup>
<b>Bioinformatics and Homology modelling</b>	Sequence alignment- Clustal Omega <sup>40</sup> , MUSCLE <sup>41</sup> , MAFFT <sup>42</sup> , COBALT <sup>43</sup> , T-COFFEE <sup>44</sup> ; Model construction- RaptorX <sup>45</sup> , MODELLER <sup>46</sup> , SWISS-MODEL <sup>47</sup>
<b>Substrate binding domain identification</b>	MetaPocket <sup>48</sup> , Site-hound <sup>49</sup> , 3DLigandSite <sup>50</sup> , CASTp <sup>51</sup> , PocketPicker <sup>52</sup> , LigAsite <sup>53</sup> , FINDSITE <sup>54</sup>
<b>Molecular graphic systems</b>	UCSF chimera <sup>32</sup> , Avogadro <sup>55</sup> , VMD <sup>56</sup> , PyMOL <sup>57</sup>
<b>High throughput/Virtual screening</b>	ZINC <sup>58</sup> , PubChem <sup>59</sup> , ZincPharmer <sup>60</sup> , ChemSpider <sup>61</sup> , DrugBank <sup>62</sup> , ChEMBL <sup>63</sup>
<b>Covalent Docking software</b>	AutoDock <sup>29</sup> , AutoDock Vina <sup>31</sup> , CovalentDock <sup>30</sup> , GOLD <sup>28,64</sup> , Glide <sup>65</sup> , CovDock <sup>10</sup> , DOCKTITE <sup>66</sup>
<b>QM/MM simulation software</b>	Amber <sup>67,68</sup> , Gromacs <sup>69</sup> , ADF <sup>70</sup> , NWChem <sup>71</sup> , CHARMM/Q-Chem <sup>72</sup> , ChemShell <sup>73</sup>

## 6 Conclusion

An overabundance of information is available on covalent HCV viral protease inhibitors however it is not able to conceal difficulties arising from the evident lack of selectivity. Covalent targeting of non-catalytic cysteine residues has transpired as an effective strategy for obtaining selectivity in HCV viral protease inhibitors. This strategy is the first of its kind to describe multiple *in silico* molecular approaches and applications that may cumulatively assist in the deliberate design of selective covalent HCV viral protease inhibitors. In the future, it is with great hope that this approach may give insights on how to design a new generation of highly selective covalent inhibitors for other protease family members and a diverse range of protein families.

**Conflicts of interest**

Authors declare no potential conflicts of interest.

**Acknowledgements**

The authors acknowledge the National Research Foundation for their financial support.

## References

- (1) Petruzzello, A., Marigliano, S., Loquercio, G., Cozzolino, A., and Cacciapuoti, C. (2016) Global epidemiology of hepatitis C virus infection: An up-date of the distribution and circulation of hepatitis C virus genotypes. *World J. Gastroenterol.* 22, 7824–7840.
- (2) Soumana, D. I., Ali, A., and Schiffer, C. A. (2014) Structural analysis of asunaprevir resistance in HCV NS3/4A protease. *ACS Chem. Biol.* 9, 2485–2490.
- (3) Manns, M. P., Buti, M., Gane, E., Pawlotsky, J. M., Razavi, H., Terrault, N., & Younossi, Z. (2017) Hepatitis C virus infection. *Nat. Rev. Dis. Prim.* 3.
- (4) Villani, R., Bellanti, F., and Serviddio, G. (2013) Treatment of Chronic HCV Infection in the Era of Protease Inhibitors, in *Practical Management of Chronic Viral Hepatitis*, pp 167–184.
- (5) Kalgutkar, A. S., and Dalvie, D. K. (2012) Drug discovery for a new generation of covalent drugs. *Expert Opin. Drug Discov.* 7, 561–581.
- (6) Kumalo, H. M., Bhakat, S., and Soliman, M. E. S. (2015) Theory and applications of covalent docking in drug discovery: Merits and pitfalls. *Molecules* 20, 1984–2000.
- (7) Adeniyi, A. A., Muthusamy, R., and Soliman, M. E. (2016) New drug design with covalent modifiers. *Expert Opin. Drug Discov.* 11, 79–90.
- (8) Awoonor-Williams, E., Walsh, A. G., & Rowley, C. N. (2017) Modeling Covalent-Modifier Drugs. *Biochim. Biophys. Acta (BBA)-Proteins Proteomics* 1–36.
- (9) Becker, D., Kaczmarek, Z., Arkona, C., Schulz, R., Tauber, C., Wolber, G., Hilgenfeld, R., Coll, M. and Rademann, J. (2016) Irreversible inhibitors of the 3C protease of Coxsackie virus through templated assembly of protein-binding fragments. *Nat. Commun.* 7, 12761.
- (10) Zhu, K., Borrelli, K. W., Greenwood, J. R., Day, T., Abel, R., Farid, R. S., and Harder, E. (2014) Docking covalent inhibitors: A parameter free approach to pose prediction and scoring. *J. Chem. Inf. Model.* 54, 1932–1940.
- (11) Hagel, M., Niu, D., St Martin, T., Sheets, M. P., Qiao, L., Bernard, H., Karp, R. M., Zhu, Z., Labenski, M. T., Chaturvedi, P., Nacht, M., Westlin, W. F., Petter, R. C., and Singh, J. (2011) Selective irreversible inhibition of a protease by targeting a noncatalytic cysteine. *Nat. Chem. Biol.* 7, 22–24.
- (12) Yao, N., Reichert, P., Taremi, S. S., Prosise, W. W., and Weber, P. C. (1999) Molecular views of viral polyprotein processing revealed by the crystal structure of the hepatitis C virus bifunctional protease-helicase. *Structure* 7, 1353–1363.

- (13) Shiryayev, S. A., Thomsen, E. R., Cieplak, P., Chudin, E., Cheltsov, A. V., Chee, M. S., Kozlov, I. A., and Strongin, A. Y. (2012) New details of HCV NS3/4A proteinase functionality revealed by a high-throughput cleavage assay. *PLoS One* 7, 1–12.
- (14) Hamad, H. A., Thurston, J., Teague, T., Ackad, E., and Yousef, M. S. (2016) The NS4A Cofactor Dependent Enhancement of HCV NS3 Protease Activity Correlates with a 4D Geometrical Measure of the Catalytic Triad Region. *PLoS One* 11, e0168002.
- (15) Chevaliez, S., and Pawlotsky, J.-M. (2006) HCV Genome and Life Cycle. *Hepat. C Viruses Genomes Mol. Biol.* 5–47.
- (16) Kim, J. L., Morgenstern, K. A., Lin, C., Fox, T., Dwyer, M. D., Landro, J. A., Chambers, S. P., Markland, W., Lepre, C. A., O'Malley, E. T., Harbeson, S. L., Rice, C. M., Murcko, M. A., Caron, P. R., and Thomson, J. A. (1996) Crystal structure of the hepatitis C virus NS3 protease domain complexed with a synthetic NS4A cofactor peptide. *Cell* 87, 343–355.
- (17) Tanoury, G. J., Eastham, S., Harrison, C. L., Littler, B. J., Ruggiero, P. L., Ye, Z., and Grillo, A. (2016) Telaprevir: From Drug Discovery to the Manufacture of Drug Substance. *Compr. Accounts Pharm. Res. Dev. From Discov. to Late-Stage Process Dev.*
- (18) Steinkühler, C., Biasiol, G., Brunetti, M., Urbani, A., Koch, U., Cortese, R., Pessi, A., and De Francesco, R. (1998) Product inhibition of the hepatitis C virus NS3 protease. *Biochemistry* 37, 8899–8905.
- (19) Lamarre, D., Anderson, P. C., Bailey, M., Beaulieu, P., Bolger, G., Bonneau, P., Bös, M., Cameron, D. R., Cartier, M., Cordingley, M. G., Faucher, A.-M., Goudreau, N., Kawai, S. H., Kukulj, G., Lagacé, L., LaPlante, S. R., Narjes, H., Poupard, M.-A., Rancourt, J., Sentjens, R. E., St. George, R., Simoneau, B., Steinmann, G., Thibeault, D., Tsantrizos, Y. S., Weldon, S. M., Yong, C.-L., and Llinàs-Brunet, M. (2003) An NS3 protease inhibitor with antiviral effects in humans infected with hepatitis C virus. *Nature* 426.
- (20) Chen, K. X., and Njoroge, F. G. (2011) NS3 Protease Covalent Inhibitors, in *Hepatitis C: Antiviral drug discovery and development.* (Tan, S.-L., and He, Y., Eds.).
- (21) Baillie, T. A. (2016) Targeted Covalent Inhibitors for Drug Design. *Angew. Chemie - Int. Ed.*
- (22) Swinney, D. C. (2004) Opinion: Biochemical mechanisms of drug action: what does it take for success? *Nat. Rev. Drug Discov.* 3, 801–808.
- (23) Johnson, D. S., Weerapana, E., and Cravatt, B. F. (2010) Strategies for discovering and derisking covalent, irreversible enzyme inhibitors. *Future Med. Chem.* 2, 949–964.



- (24) Johansson, M. H. (2012) Reversible Michael additions: covalent inhibitors and prodrugs. *Mini Rev Med Chem* 12, 1330–1344.
- (25) Hallenbeck, K. K., Turner, D. M., Renslo, A. R., and Arkin, M. R. (2017) Targeting Non-Catalytic Cysteine Residues Through Structure-Guided Drug Discovery. *Curr. Top. Med. Chem.* 17, 4–15.
- (26) Blake, L., and Soliman, M. E. S. (2014) Identification of irreversible protein splicing inhibitors as potential anti-TB drugs: Insight from hybrid non-covalent/covalent docking virtual screening and molecular dynamics simulations. *Med. Chem. Res.* 23, 2312–2323.
- (27) Ward, R. A., Colclough, N., Challinor, M., Debreczeni, J., Eckersley, K., Fairley, G., Feron, L., Flemington, V., Graham, M. A., Greenwood, R., Hopcroft, P., Howard, T. D., James, M., Jones, C. D., Jones, C. R., Renshaw, J., Roberts, K., Snow, L., Tonge, M., and Yeung, K. (2015) Structure-guided Design of Highly Selective and Potent Covalent Inhibitors of ERK1 / 2 58, 4790–4801.
- (28) Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W., and Taylor, R. D. (2003) Improved protein-ligand docking using GOLD. *Proteins Struct. Funct. Genet.* 52, 609–623.
- (29) Cosconati, S., Forli, S., Perryman, A. L., Harris, R., Goodsell, D. S., and Olson, A. J. (2010) Virtual Screening with AutoDock: Theory and Practice. *Expert Opin. Drug Discov.* 5, 597–607.
- (30) Ouyang, X., Zhou, S., Su, C. T. T., Ge, Z., Li, R., and Kwoh, C. K. (2013) CovalentDock: Automated covalent docking with parameterized covalent linkage energy estimation and molecular geometry constraints. *J. Comput. Chem.* 34, 326–336.
- (31) Trott, O., and Olson, A. J. (2010) AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J. Comput. Chem.* 31, 455–61.
- (32) Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* 25, 1605–1612.
- (33) Xiao, C., and Zhang, Y. (2007) Design-atom approach for the quantum mechanical/molecular mechanical covalent boundary: A design-carbon atom with five valence electrons. *J. Chem. Phys.* 127, 1–25.
- (34) Allen, M. (2004) Introduction to molecular dynamics simulation. *Comput. Soft Matter From Synth. Polym. to proteins* 23, 1–28.
- (35) Groenhof, G. (2013) Introduction to QM/MM simulations., in *Biomolecular simulations: methods and protocols*, p 43–66.

- (36) Olayide A., and M.E., S. (2017) Quantum mechanics implementation in drug design workflows: Does it really help? *Drug Des. Dev. Ther. Under Rev.*
- (37) Senn, H. M., and Thiel, W. (2009) QM/MM methods for biomolecular systems. *Angew. Chemie Int. Ed.* 48, 1198–1229.
- (38) Karplus, M., and McCammon, J. A. (2002) Molecular dynamics simulations of biomolecules. *Nat. Struct. Mol. Biol.* 9, 646–652.
- (39) Rose, P. W., Bi, C., Bluhm, W. F., Christie, C. H., Dimitropoulos, D., Dutta, S., Green, R. K., Goodsell, D. S., Prlić, A., Quesada, M., Quinn, G. B., Ramos, A. G., Westbrook, J. D., Young, J., Zardecki, C., Berman, H. M., and Bourne, P. E. (2013) The RCSB Protein Data Bank: New resources for research and education. *Nucleic Acids Res.* 41, D475–D482.
- (40) Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., and Higgins, D. G. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7, 539.
- (41) Edgar, R. C. (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797.
- (42) Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30, 3059–3066.
- (43) Papadopoulos, J. S., and Agarwala, R. (2007) COBALT: Constraint-based alignment tool for multiple protein sequences. *Bioinformatics* 23, 1073–1079.
- (44) Di Tommaso, P., Moretti, S., Xenarios, I., Orobic, M., Montanyola, A., Chang, J. M., Taly, J. F., and Notredame, C. (2011) T-Coffee: A web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Res.* 39.
- (45) Källberg, M., Wang, H., Wang, S., Peng, J., Wang, Z., Lu, H., and Xu, J. (2012) Template-based protein structure modeling using the RaptorX web server. *Nat. Protoc.* 7, 1511–1522.
- (46) Fiser, A., and Šali, A. (2003) MODELLER: Generation and Refinement of Homology-Based Protein Structure Models. *Methods Enzymol.* 374, 461–491.
- (47) Arnold, K., Bordoli, L., Kopp, J., and Schwede, T. (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* 22, 195–201.
- (48) Huang, B. (2009) MetaPocket: A Meta Approach to Improve Protein Ligand Binding Site Prediction. *Omi. A J. Integr. Biol.* 13, 325–330.

- (49) Hernandez, M., Ghersi, D., and Sanchez, R. (2009) SITEHOUND-web: A server for ligand binding site identification in protein structures. *Nucleic Acids Res.* 37, W413–W416.
- (50) Wass, M. N., Kelley, L. A., and Sternberg, M. J. E. (2010) 3DLigandSite: Predicting ligand-binding sites using similar structures. *Nucleic Acids Res.* 38, gkq406.
- (51) Dundas, J., Ouyang, Z., Tseng, J., Binkowski, A., Turpaz, Y., and Liang, J. (2006) CASTp: Computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Res.* 34, 116–118.
- (52) Weisel, M., Proschak, E., and Schneider, G. (2007) PocketPicker: analysis of ligand binding-sites with shape descriptors. *Chem. Cent. J.* 1, 7.
- (53) Dessailly, B. H., Lensink, M. F., Orengo, C. A., and Wodak, S. J. (2008) LigASite - A database of biologically relevant binding sites in proteins with known apo-structures. *Nucleic Acids Res.* 36.
- (54) Brylinski, M., and Skolnick, J. (2008) A threading-based method (FINDSITE) for ligand-binding site prediction and functional annotation. *Proc. Natl. Acad. Sci. U. S. A.* 105, 129–134.
- (55) Hanwell, M. D., Curtis, D. E., Lonie, D. C., Vandermeersch, T., Zurek, E., and Hutchison, G. R. (2012) Avogadro: An advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminform.* 4, 17.
- (56) Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: Visual molecular dynamics. *J. Mol. Graph.* 14, 33–38.
- (57) DeLano, W. L. (2014) The PyMOL Molecular Graphics System, Version 1.8. Schrödinger LLC <http://www.pymol.org>.
- (58) Irwin, J. J., Sterling, T., Mysinger, M. M., Bolstad, E. S., and Coleman, R. G. (2012) ZINC: A free tool to discover chemistry for biology. *J. Chem. Inf. Model.*
- (59) Xie, X.-Q. S. (2010) Exploiting PubChem for virtual screening. *Expert Opin. Drug Discov.* 5, 1205–1220.
- (60) Koes, D. R., and Camacho, C. J. (2012) ZINCPharmer: Pharmacophore search of the ZINC database. *Nucleic Acids Res.* 40.
- (61) Pence, H. E., and Williams, A. (2010) Chemspider: An online chemical information resource. *J. Chem. Educ.*
- (62) Wishart, D. S. (2006) DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.* 34, D668–D672.

- (63) Gaulton, A., Bellis, L. J., Bento, A. P., Chambers, J., Davies, M., Hersey, A., Light, Y., McGlinchey, S., Michalovich, D., Al-Lazikani, B., and Overington, J. P. (2012) ChEMBL: A large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* 40.
- (64) Jones, G., Willett, P., Glen, R. C., Leach, A. R., and Taylor, R. (1997) Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* 267, 727–48.
- (65) Friesner, R. A., Murphy, R. B., Repasky, M. P., Frye, L. L., Greenwood, J. R., Halgren, T. A., Sanschagrin, P. C., and Mainz, D. T. (2006) Extra precision glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J. Med. Chem.* 49, 6177–6196.
- (66) Schmidt, B. (2014) DOCKTITE - A Highly Versatile Step-by-Step Workflow for Covalent Docking and Virtual Screening in MOE. *J. Chem. Inf. Model.* 141226132628000.
- (67) Götz, A. W., Clark, M. A., and Walker, R. C. (2014) An extensible interface for QM/MM molecular dynamics simulations with AMBER. *J. Comput. Chem.* 35, 95–108.
- (68) Salomon-Ferrer, R., Case, D. A., and Walker, R. C. (2013) An overview of the Amber biomolecular simulation package. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 3, 198–210.
- (69) Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., and Berendsen, H. J. C. (2005) GROMACS: Fast, flexible, and free. *J. Comput. Chem.*
- (70) Velde, G. T. E., Bickelhaupt, F. M., Baerends, E. J., Guerra, C. F., and Gisbergen, S. J. A. V. A. N. (2001) Chemistry with ADF 22, 931–967.
- (71) Valiev, M., Bylaska, E. J., Govind, N., Kowalski, K., Straatsma, T. P., Van Dam, H. J. J., Wang, D., Nieplocha, J., Apra, E., Windus, T. L., and De Jong, W. A. (2010) NWChem: A comprehensive and scalable open-source solution for large scale molecular simulations. *Comput. Phys. Commun.* 181, 1477–1489.
- (72) Woodcock, H. L., Hodošček, M., Gilbert, A. T. B., Gill, P. M. W., Schaefer, H. F., and Brooks, B. R. (2007) Interfacing Q-Chem and CHARMM to perform QM/MM reaction path calculations. *J. Comput. Chem.* 28, 1485–1502.
- (73) Metz, S., Kästner, J., Sokol, A. A., Keal, T. W., and Sherwood, P. (2014) ChemShell-a modular software package for QM/MM simulations. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 4, 101–110.

## CHAPTER 5

### **Targeting HCV Polymerase: A Structural and Dynamic Perspective into the Mechanism of Selective Covalent Inhibition**

Letitia Shunmugam<sup>a</sup> and Mahmoud E. S. Soliman<sup>a\*</sup>

<sup>a</sup> Molecular Bio-Computation and Drug Design laboratory, School of Health Sciences, University of KwaZulu-Natal, Westville Campus, Durban 4001, South Africa.

\*Corresponding Author: Mahmoud E.S. Soliman

School of Health Sciences, University of KwaZulu-Natal, Westville Campus, Durban 4001, South Africa

Email: [soliman@ukzn.ac.za](mailto:soliman@ukzn.ac.za)

Telephone: +27 (0) 31 260 8048, Fax: +27 (0) 31 260 7872

## **Abstract**

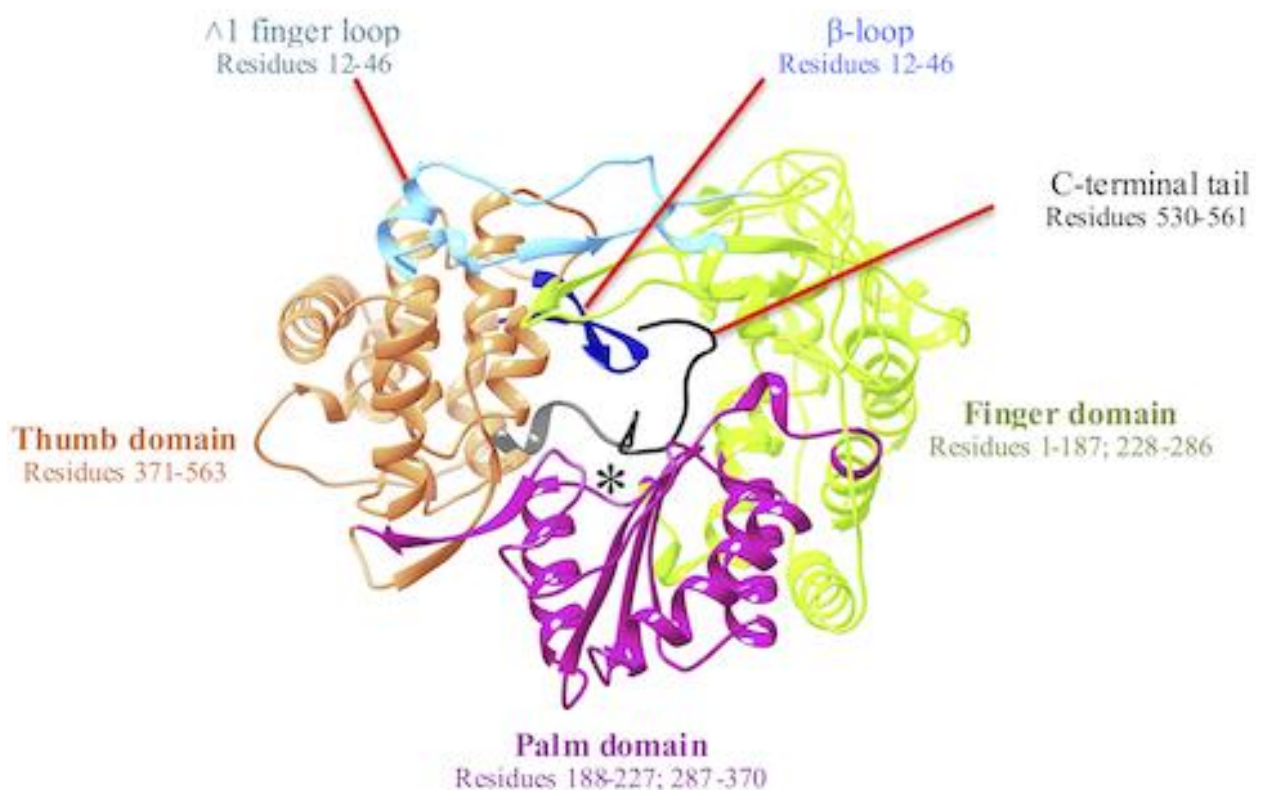
Background: Hepatitis C Virus (HCV) has raised alerting concerns over its emerging pandemic status. Current market-available drugs lack specificity, stability and potency. The HCV NS5B RNA-dependent RNA polymerase (RdRp) is a vital component in viral replication and is often targeted in antiviral therapies. Recent experimental procedures have led to the discovery of a novel covalent RdRp inhibitor, compound 47, which selectively targets cysteine 366 of HCV RdRp and exhibits promising pharmacokinetic outcomes. Selective covalent inhibition of HCV is, however, a highly neglected domain in literature that is intensified by the evident lack of efficient structure-based drug design protocols. Herein, we provide an atomistic insight into a novel selective approach to inhibit HCV RdRp. Methodology/Results: Covalent molecular dynamic analyses revealed the inhibitory mechanism of compound 47 on the RdRp. Inhibitor binding induced distinctive internal movements resulting in the disruption of normal physiological interdomain interactions. Conclusion: Compound 47 stimulates reorganization of key protein elements required for RNA transcription, thus hampering viral replication as well disrupting the overall conformation of HCV. This study will open a new avenue for the design of novel selective inhibitors against HCV as well as other viral families.

**Keywords:** Hepatitis C virus, polymerase, selective covalent inhibition, antiviral research

## 1 Introduction

Hepatitis C virus (HCV) is a membrane-bound, hepatotropic RNA disease <sup>1</sup>. Over the years, proceeding its initial discovery in 1989 <sup>2</sup>, HCV has garnered global concern over its pandemic ascent. It has been approximated that at least 3% of the world's population have been infected with HCV, of which a further 30% will go on to develop severe hepatic-related diseases including cirrhosis and hepatocellular cancers <sup>3</sup>.

The HCV genome is translated into a polymer entailing 3010 amino-acid residues, which are cleaved by both viral and cellular proteases to generate proteins necessary for replication and viral assembly <sup>4,5</sup>. Amid these proteins, is the non-structural 5B RNA-dependent RNA polymerase (NS5B RdRp). The RdRp of HCV is a 66kDa protein comprised of approximately 591 amino acids and is located at the C-terminal of the viral genome. The RdRp occupies a distinguishable three-dimensional structure that embraces a right-handed topology with distinctive domains recognized universally as the finger, palm, and thumb regions <sup>6</sup> (Figure 1).



**Figure 1:** Graphical representation of right-hand x-ray crystallography structure of NS5B RNA dependent RNA polymerase (PDB ID: 3H5S). The active site is denoted by \*.

Identified as an essential component of the HCV viral life cycle, RdRp adopts a vital role in facilitation of RNA genome replication and transcription. The RdRp has two modes of enzyme activity that corresponds to the different conformations. Extensive interdomain contact between the finger and thumb domains facilitate the enzyme's closed "active" conformation, which permits the *de novo* synthesis of RNA for initiation of viral replication. On the other hand, the enzyme adopts a more open conformation by displacement of the C-terminal tail that also serves as the linker region and the  $\beta$ -loop. In literature, the open conformation of RdRp is still an ongoing investigation and very little is understood at this moment however evidence seems to suggest that this state is an indication of enzyme inactivity as intramolecular domain interactions vital for the initiation of replication is disturbed <sup>7</sup>. Thus, RdRp and its conformation serves immense importance in viral replication, for this reason, it is frequently targeted by the medicinal chemistry and pharmaceutical industry for the design and development of potent antiviral therapies <sup>8</sup>.


Over the last twenty years, several drug candidates have been discovered and implemented in HCV therapeutics. The drugs are classified into two distinct classes: nucleoside/nucleotide inhibitors (NIs) and non-nucleoside inhibitors (NNIs), categorization is dependent on the mechanism of action <sup>9</sup>. The NIs serves as imitator substrates for the RdRp to avert RNA chain elongation through nucleoside triphosphate displacement <sup>10</sup>. Notable NIs include valopicitabine (Indenix), Balapiravir (Roche), Mericitabine (Pharmasset/ Roche) <sup>11,12</sup> and Uprifosbuvir (Merck & Co) <sup>13,14</sup>. Contrastingly, NNIs preferably bind to allosteric sites in the RdRp palm or thumb domains, initiating conformational modifications of the enzyme, which hinders its function in the initiation of RNA synthesis <sup>10</sup>. Examples of NNIs include ABT-072, setrobuvir and dasabuvir <sup>15,16</sup>.

Although the current market available HCV RdRp inhibitors exhibit potent activity against the viral enzyme, their poor chemical stability and relatively high molecular weights have led to challenges arising in pharmacokinetics especially concerning drug absorption, distribution, biotransformation and elimination <sup>17,18</sup>. Furthermore, the evolution and adaption of HCV in favour of survival has primed the onset of mutations at numerous sites leading to genetic variability amongst genotypes. A process that's further escalated by RdRp's lack of adequate proofreading capabilities hence the enzyme is unable to correct mechanistic errors, this may potentially lead to multiple genetic variations of the virus within a single infected individual <sup>19</sup>. The generation of HCV mutant forms have actively endorsed the resistance of the virus to non-selective enzyme inhibitors <sup>20,21</sup>. In due course, researchers developed a new therapeutic approach that specifically targeted RdRp of multiple if not all HCV geno- and sub-type, later signified as selective covalent inhibition. In theory, the approach appeared to fulfill all aspects



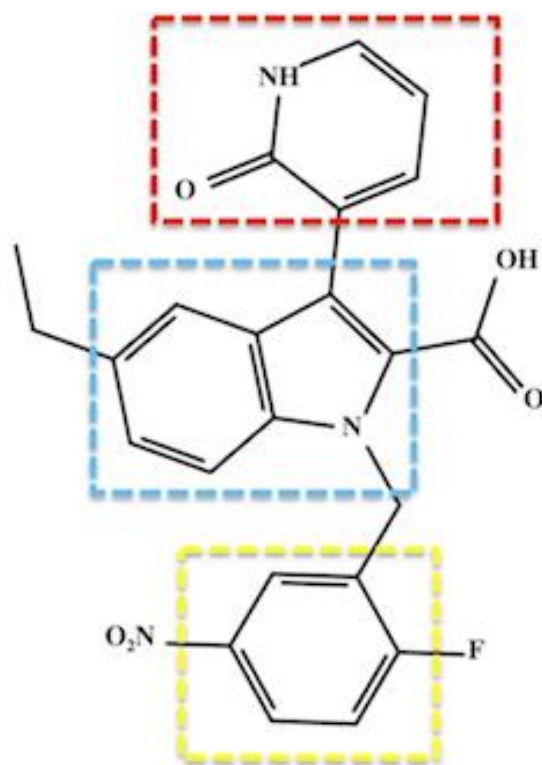
lacking in current therapies however its practicality was certainly questioned due to difficulties encountered during its implementation in drug discovery as a result of scarce availability of competent protocols <sup>22</sup>.

Selective covalent inhibition involves the binding of an inhibitor possessing a highly selective nature, to a residue belonging to a serine or cysteine amino acid in the protein of interest. The designated amino acid can be extrapolated from literature whereby experimental investigations reveal its conservative nature across multiple genotypes of a virus. The resultant bond formed between the two components are well-thought-out as an irreversible drug-protein interaction <sup>23</sup>. In 2006, Lee and co-investigators were the first to identify the importance of cysteine 366 (Cys366) in specific targeting of HCV RdRp <sup>24</sup>. The study discovered that mutation of cysteine to glycine (C366G) results in drastic loss of wild-type RdRp enzyme activity. These findings profoundly disclosed the importance of Cys366 in selective inhibitor binding and full enzymatic functioning of HCV RdRp. The Cys366 amino acid residue is conserved across all HCV RdRp sequences known to date (Figure 2). Thus rendering Cys366 as an ideal target for obtaining selectivity in covalent HCV inhibition <sup>24</sup>. However selective covalent inhibition in association with HCV RdRp has not yet been adequately investigated, which is evident by the cavernous lack of subsistence within literature. This breach in science has negatively impacted the design and discovery of potentially effective HCV antiviral therapies <sup>25,26</sup>.

**1a** QPEYDLELITS**C**SSNVSVAH  
**1b** QPEYDLELITS**C**SSNVSVAH  
**1c** QPEYDLELITS**C**SSNVSVAH  
**2a** RPEYDLELITS**C**SSNVSVAH  
**2b** RPEYDLELITS**C**SSNVSVAH  
**2c** KAEYDLELITS**C**SSNVSVAH  
**3a** QPTYDLELITS**C**SSNVSVAR  
**3b** QPTYDLELITS**C**SSNVSVAC  
**4a** QPAYDLELITS**C**SSNVSVAH  
**5a** VPAYDLELVTS**C**SSNVSVAR  
**6a** QPTYDLELITS**C**SSNVSVAH  
**6b** QPTYDLELITS**C**SSNVSVAH  
**6c** HPEYDLELITS**C**SSNVSVAH  
**6l** RPEYDLELITS**C**SSNVSVAH  
**6o** HPEYDLELITS**C**SSNVSVAH  
**6p** RPEYDLELITS**C**SSNVSVAH  


**Figure 2:** Sequence alignment of NS5B polymerase. Cysteine 366 is conserved within all genotypes of HCV. C: cysteine (red); Yellow star indicates identical residues. The genome sequences for each of the HCV genotypes were obtained from UniProt and subsequently aligned using ClustalW2 <sup>27,28</sup>.

Soon after its discovery, Chen *et al* (2012), issued substantial findings supporting the concept of selective covalent inhibition and specific targeting of RdRp Cys366 <sup>6</sup>. The study involved synthesis of inhibitors as part of structure-activity relationship optimization (SAR) program, this sequentially led to the discovery of 3-(1,2-Dihydro-2-oxo-3-pyridinyl)-5-ethyl-1-[(2-fluoro-5-nitrophenyl)methyl]-1H-indole-2-carboxylic acid, better referred to as compound 47 (Figure 3). The covalent inhibitor's global experimental profile revealed propitious *in vitro* safety, PK and potency. To date, there is no experimental data investigating the structural dynamic entities regarding compound 47 and RdRp interactions.



**Figure 3:** Two-dimensional structural representation of HCV RdRp covalent inhibitor, compound 47. Red: C-3 pyridone ring; Blue: indole core and Yellow: benzene ring. The compound exhibited covalent binding to RdRp whereby the thiol group of Cys366 attacked the benzene ring at the position para to the nitro group and the fluoro group is released presumably through an aromatic nucleophilic substitution reaction <sup>6</sup>.

In this study, we investigate for the first time, selective covalent inhibition of HCV RdRp. Through extensive application of various molecular and bioinformatics tools, we expressed and compared RdRp dynamic and structural characteristics of the free enzyme as well as a covalently-bound complex. This study will emphasize the importance and necessity of selective covalent inhibition in antiviral therapy. It is with great hope that this study will aid in the design and development of potent, target-specific covalent inhibitors against HCV as well as a wide variety of other viruses.

## 2 Computational Methodology

### 2.1 System preparation

The HCV RdRp and the three-dimensional structure of the experimental HCV inhibitor, compound 47, were obtained from RSCB Protein Data Bank (PDB code: 3H5S and 3TYQ, respectively) <sup>6,29</sup>. The apo system contains only the free enzyme. The bound system's protonation states was optimized, directly followed by the addition of hydrogen atoms through Protein Preparation Wizard in Maestro <sup>30,31</sup>.

Molecular docking was utilized to envisage binding affinities and optimized conformations of compound 47 within the RdRp active site <sup>32</sup> (grid box spacing of 0.375 Å and x, y, z dimensions of 9 x 9 x 7.72). Docking software utilized in this study included AutoDock Vina <sup>33</sup> and UCSF Chimera <sup>34</sup>. Compound 47 was subsequently docked in the active binding site of RdRp and the resultant complex accompanied by the most negative binding energy (kcal.mol<sup>-1</sup>) was selected and subjected to molecular dynamic simulations.

## 2.2 Molecular Dynamics Simulation

Molecular dynamic (MD) simulations provide a powerful tool aimed at the exploration of the physical activity perpetrated by molecules and atoms. Thus providing a focused perspective into the biomolecular processes of biological systems <sup>35</sup>. The Amber14 package <sup>36,37</sup> was used to perform a 200 nanoseconds (ns) MD simulation for the apo and bound system. The MD simulation was performed using the GPU version of the PMEMD engine provided through Amber and the FF14SB Amber force field. The antechamber module was utilised to generate partial atomic charges for the compound 47 by applying General Amber Force Field (GAFF) and restrained electrostatic potential (RESP) procedures <sup>38</sup>. The LEaP module was utilised to neutralise and solvate the both the apo and bound system systems through the addition of hydrogen atoms, sodium and chloride ions. Owing to the blatant lack of effective protocols within literature, for efficient selective covalent inhibition, we devised an in-house protocol to enhance *in silico* outcomes of selective covalent inhibitors. A greater expansion of the mentioned protocol can be found in the referenced articles <sup>39,40</sup>. The covalent system topology and input coordinate files were created using Dabble <sup>41</sup>. To begin with, the covalently bound system was minimized in two steps with the restraint potential of 10 Å for consideration of solute molecule using 500 steepest descent steps, followed by 1000 steps of conjugate gradient minimization. The system was then gradually heated from 0 to 300 Kelvin (K) and 2.5ns of equilibration was executed for apo and bound system stabilisation. Potential harmonic restraint of 10 kcal.mol<sup>-1</sup> Å<sup>2</sup> for solute atoms and Langevin thermostat with collision frequency of 1 ps was applied to the system. The pressure and the number of atoms were kept constant to bear resemblance to an isobaric–isothermal ensemble (NPT). The pressure of both the systems were maintained at 1 bar using the Berendsen barostat <sup>42</sup> and the SHAKE algorithm was utilised for hydrogen bond constraint <sup>43</sup>.

## 2.3 Post analysis

The coordinates of the apo and bound systems were each saved every 1 ps for which the trajectories were analysed using CPPTRAJ and PTRAJ modules in the Amber14 package. Analysis was conducted for the following: root mean square deviation (RMSD), root mean square fluctuations, radius of gyration, dynamics cross correlation matrix, template channel width, interdomain angle and solvent

accessible surface analysis (SASA). Visualisation and graphical software programs were utilised throughout this study including visual molecular dynamics (VMD)<sup>44</sup> to visualize projected trajectories; Maestro<sup>31</sup> software for preparation of the systems and lastly Origin software (OriginLab, Northampton, MA) for the plotting and generation of graphs.

### 2.3.1 Root mean square fluctuation (RMSF)

The fluctuation of individual enzyme residues regarding its average position within a given MD simulation is referred to the root mean square fluctuation (RMSF)<sup>45</sup>. The RMSF analysis provides insights into the flexibility of various regions of the HCV RdRp upon binding of the ligand and is mathematically calculated as follows:

$$sRMSF_i = \frac{(RMSF_i - RMSF)}{\sigma(RMSF)}$$

RMSF<sub>i</sub> represents the RMSF of the *i*th residue, from which the average RMSF is subtracted. This is then divided by the RMSF's standard deviation [ $\sigma(RMSF)$ ] to yield the resultant standardised RMSF (sRMSF<sub>i</sub>).

### 2.3.2 Radius of Gyration (RoG)

The RoG describes the root mean square distance of the atoms from their common center of gravity of a given enzyme molecule. This method of analysis allows the assessment of protein compactness along a given trajectory in a MD simulation. The equation below describes how RoG is estimated:

$$r^2g = \frac{\sum_{i=1}^n w_i (r_i - r^-)^2}{\sum_{i=1}^n w_i}$$

Where:  $r_i$  is the position of the *i*th atom and  $r^-$  is the center mass of atom *i*. The mean value is calculated by taking the RoG values over the number of frames in a given trajectory<sup>46</sup>.

### 2.3.3 Dynamic cross-correlation matrix (DCCM) analysis

Dynamic cross correlation is a method that's widely employed in MD simulations to effectively quantify the correlation co-efficient of motion between the atoms of a protein. The dynamic cross correlation between the residue-based fluctuations during simulation was calculated using the CPPTRAJ module incorporated in AMBER 14. The formula used to describe dynamic cross correlation is given below:

$$C_{ij} = \frac{\langle \Delta r_i \Delta r_j \rangle}{(\langle \Delta r_i^2 \rangle \langle \Delta r_j^2 \rangle)^{\frac{1}{2}}}$$

The abbreviated terms denote the following:  $C_{ij}$ : cross-correlation coefficient, fully correlated (-1) to anti-correlated (+1);  $i$ :  $i^{\text{th}}$  residue;  $j$ :  $j^{\text{th}}$  residue;  $\Delta r_i$ : displacement vectors correspond to  $i^{\text{th}}$  residue and  $\Delta r_j$ : displacement vectors correspond to  $j^{\text{th}}$  residue. The resultant DCCM of each system was constructed using Origin software.

### 2.3.4 Thermodynamic Binding free energy calculations

Molecular dynamic simulations allow admittance to free energy differences which controls the underlying mechanism of all biological processes <sup>47</sup>. Calculations for binding free energy (BFE) is an imperative method utilised to observe the in-depth binding mechanism between a protein and ligand, all-encompassing both entropic and enthalpic contributions <sup>48</sup>. Estimation of the binding affinity within a docked system is calculated through BFE using the Molecular Mechanics/GB Surface Area (MM/GBSA) technique. This method can be utilized to provide reproducible relative binding affinities of compounds with respectable accuracy and requires substantially less computational resources in comparison to a full-scale MD free energy perturbation/thermodynamic simulation <sup>49</sup>. In this study, BFE was averaged over 20 000 snapshots, which were generated from the 200ns trajectory. The explicit solvent employed in the MD simulation was discarded and substituted with a dielectric continuum as per MMGBSA protocol <sup>50</sup>. Changes in each term between the apo and bound states were calculated and contributed to the total relative BFE <sup>49,50</sup>. Molecular mechanics force fields were then employed to calculate energy contributions from the atomic coordinates of the ligand, receptor and complex in a gaseous phase. The equations below elaborate the process in which the binding free energies ( $\Delta G$ ) were assessed:

$$(1) \Delta G_{bind} = \Delta G_{complex} - \Delta G_{receptor} - \Delta G_{ligand}$$

$$(2) \Delta G_{bind} = E_{gas} + G_{sol} - T\Delta S$$

$$(3) E_{gas} = E_{int} + E_{vdw} + E_{ele}$$

$$(4) G_{sol} = G_{GB} + G_{SA}$$

$$(5) G_{SA} = \gamma SASA$$

Where the following terms are referred to:  $\Delta G_{bind}$  is gas-phase summation, gas-phase energy ( $E_{gas}$ ) and the solvation energy,  $G_{sol}$ , is less the entropy ( $T\Delta S$ ) term. The  $E_{gas}$  is the sum of internal energy,  $E_{int}$ , van der Waals (vdW) energy,  $E_{vdw}$  and electrostatic energy,  $E_{ele}$ . The total solvation energy is calculated by a summation of the total energy contributions of polar and non-polar states ( $G_{GB}$  and  $G_{SA}$ , respectively). The  $G_{SA}$  is calculated by means of the solvent accessible surface area, generated by water probe radius of 1.4 Å. Resolution of  $G_{GB}$  equation allows the determination for the energy contributions

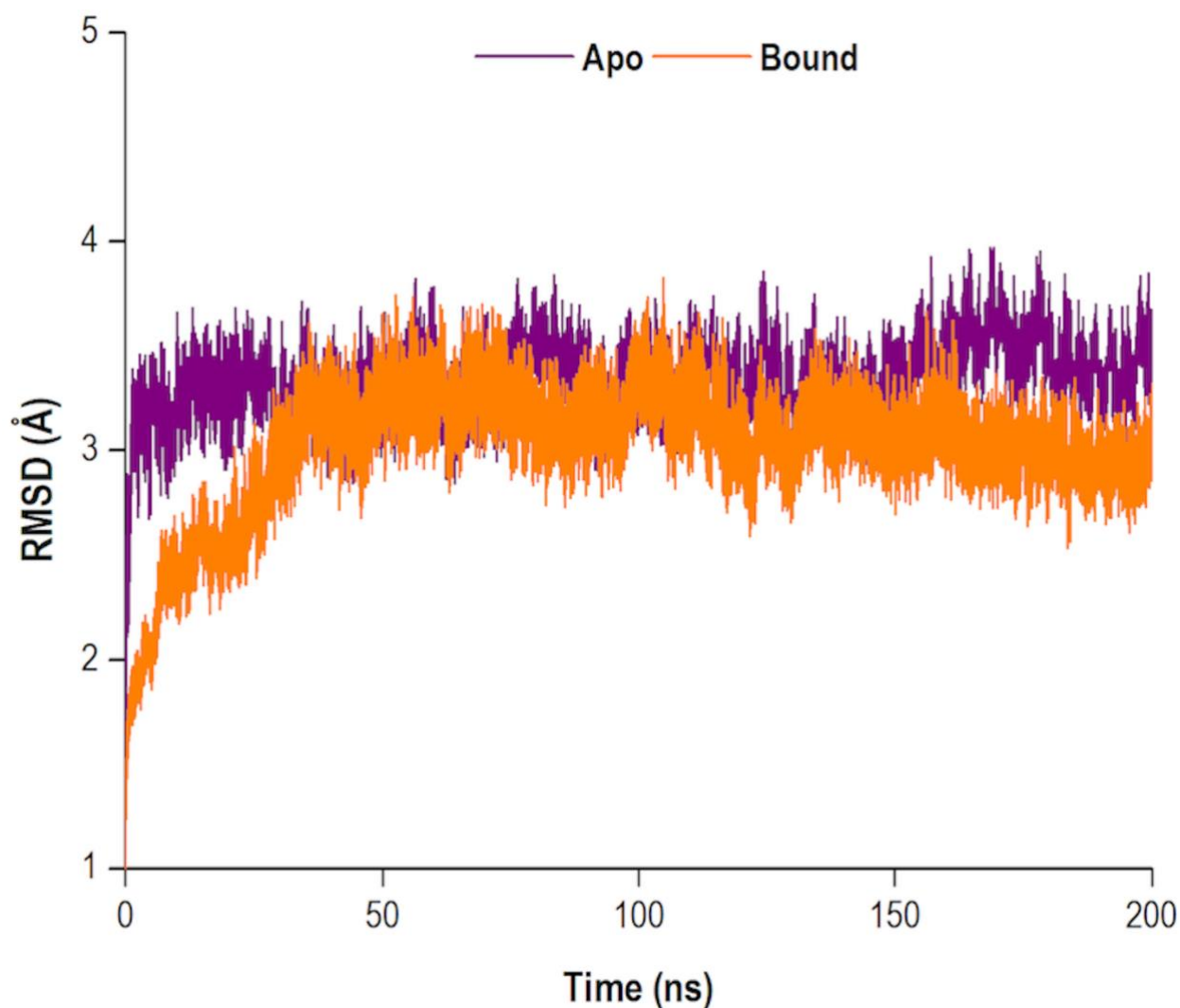
of polar states. The total entropy of the solute is denoted as ‘S’ and temperature as ‘T’. The solute and solvent dielectric constants are set to 1 and 80, respectively <sup>51,52</sup>.

The MM/GBSA method was also used to calculate final energy per-residue decomposition <sup>53</sup>. Regarding the estimated relative binding free energies, the degree of accuracy may be enhanced if the terms in the equation, particularly those in Equation 2, are averaged over several MD snapshots but typically depends on the research of interest <sup>50</sup>. Conduction of separate MD simulations for the receptor, ligand and complex will produce more accurate BFE results, however it requires greater computational resources which were not readily available for this study <sup>49,50</sup>. Though the MM/GBSA technique lacks the required accuracy for absolute BFE estimates <sup>51,54</sup>, several of our previous reports including studies from our research group have successfully employed the MM/GBSA approach in obtaining relative BFE for MD simulated systems which have been validated with experimental findings <sup>38,49,55–59</sup>. To obtain individual residue contribution to the total BFE profile of compound 47 with the HCV RdRp, per-residue free energy decomposition was performed at an atomic level for all-important residues using the MM/GBSA method using Amber14.

### **3 Results and Discussion**

#### **3.1 Stability of NS5B RdRp apo and bound system.**

Molecular dynamic (MD) simulations yield comprehensive insights into the structure and dynamics of a biological system. The convergence of a system may be reached as the system gains stability, often occurring in an increasingly longer MD simulation. Therefore, the structural dynamic examination of a system is highly dependent on the timescale of the MD simulation <sup>60</sup>. Throughout the 200ns simulation, the root mean square deviation (RMSD) of the C- $\alpha$  (alpha carbon) backbone were observed for both the apo and bound systems. The RMSD analysis was used to graphically monitor system convergence <sup>61</sup>. Both the apo and bound systems converged at approximately 30 and 35ns, respectively (Figure 4).

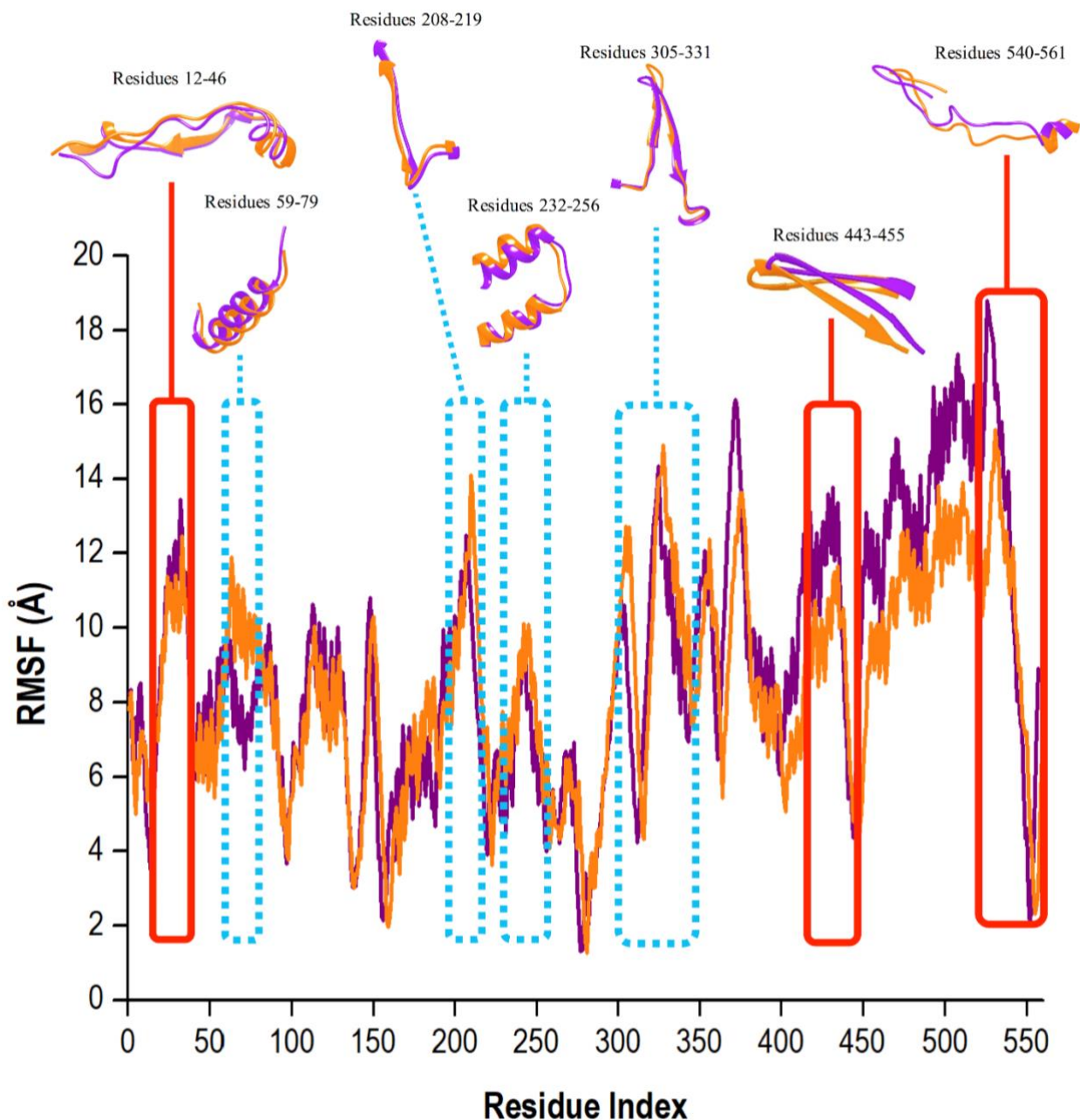


**Figure 4:** RMSD plot of C- $\alpha$  atoms of the apo and bound systems.

### 3.2 Conformational fluctuations of NS5B RdRp

To determine whether binding of compound 47 affects residue dynamic behavior, the root mean square fluctuation (RMSF) values of the apo and bound systems were analysed. The RMSF with respect to the averaged MD simulation conformation is utilized as a mean to demonstrate the differences arising in the flexibility of the residues. The RMSF of all residues were calculated for structure backbone flexibility and was represented as Figure 5.





**Figure 5:** Residue-based average C- $\alpha$  fluctuations of the apo and bound conformation of HCV RdRp throughout 200ns MD simulation.

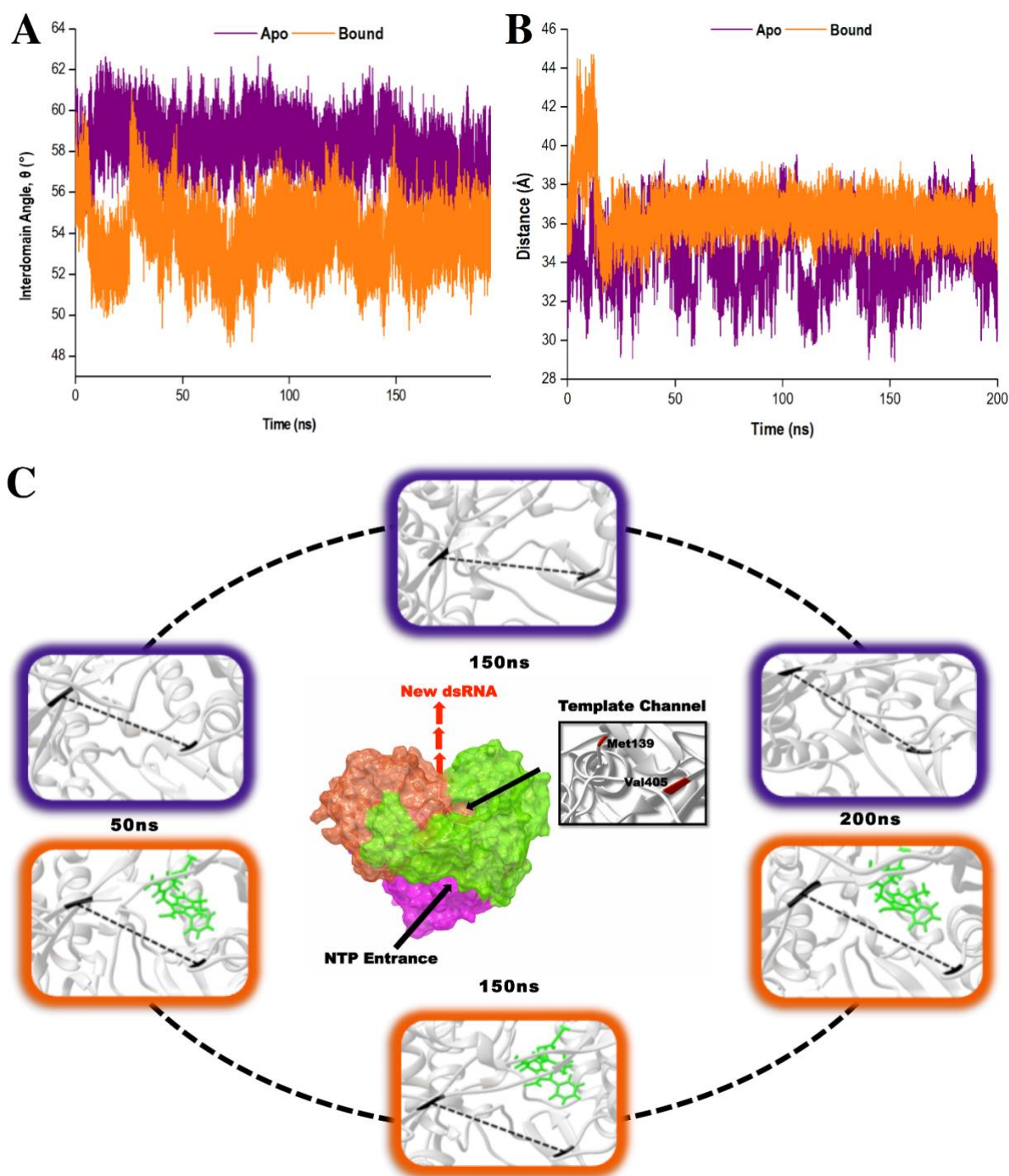
Higher C- $\alpha$  fluctuation values signify more flexible movements in relation to the average position of the residues, contrastingly, reduced values express restricted fluctuations during the simulation <sup>62</sup>. In Figure 5, the apo RdRp presented with an overall higher residual fluctuation (8.97 Å) when compared to the bound system (8.47 Å). This degree of high flexibility is concurrent with literature whereby this property enables the enzyme to accept single-stranded RNA templates, in doing so; the enzyme can effectively propagate viral replication. This finding strongly indicates that binding of compound 47 to

the RdRp, in effect, lowered dynamic residual fluctuations of the enzyme thus inducing stability of the bound state.

To effectively understand a protein's function at an atomic perspective, experimental analysis of the structure is required. It is important to note that the functional properties of a protein are determined not only by its relative structural rigidity but also by its dynamic behaviors<sup>63 64</sup>. The binding site residues of compound 47 within 5Å: Arg158, Phe193, Pro197, Arg200, Asn316, Asp318, Asp319, Thr364, Ser365, Ser367, Ser368, Leu384, Arg386, Arg399, Ser407, Gly410, Asn411, Met414, Tyr415, Gln446, Ile447, Tyr448, Gly449, Ala450, Tyr555 and Ser556. Compound 47 is bound covalently to Cys366 located in the palm domain of the RdRp within the active binding site. In Figure 5, close observation revealed that the unbound RdRp experienced a great degree of flexibility as regions assume differing conformations in accordance with opening and closing of the active binding site. Upon binding of compound 47, the residual fluctuations of RdRp are lowered and regions greatly affected by the covalent binding encompass residues 12-46, 443-455 and 530-561. During the 200ns MD simulation, the covalently bound system did experience instances of elevated C- $\alpha$  fluctuations. The regions with the most notable fluctuations were residues 59-79, 208-219, 232-256 and 305-331 (Figure 5, highlighted in blue). This is owed to RdRp undergoing conformational adjustments in order to accommodate compound 47 binding and subsequent ligand-enzyme interactions. In essence, compound 47 caused a certain degree of interference with the enzyme's conformation, which is dire for optimum enzyme functionalities, leading to interruptions in or possibly extermination of downstream activities important in viral replication<sup>8,16</sup>.

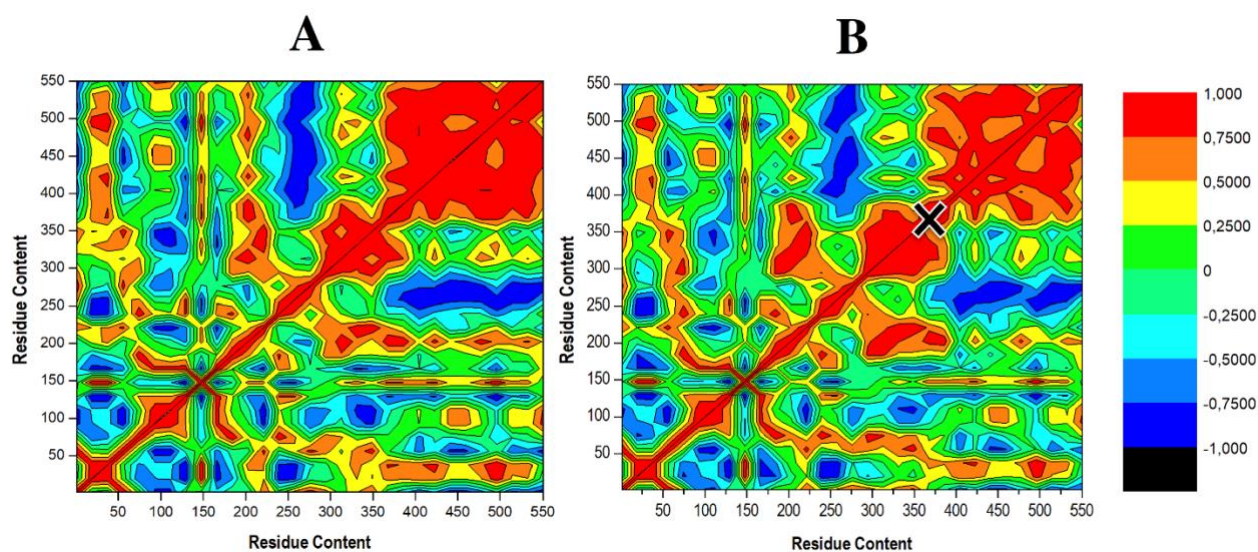
To provide further insight into the impact of compound 47 binding on conformational sampling, template channel width (TCW) and interdomain angles were investigated. Narrower widths and smaller angles correspond to a more closed enzyme conformation<sup>65</sup>. In Figure 6A and B, we observed that the apo system displayed an average interdomain angle of 58.92° and experienced fluctuations in TCW ranging between 28.15Å and 39.55Å. Contrastingly, the covalently bound system exhibited a more compact structure with an average interdomain angle of 53.92° and a narrower TCW range of 33.39Å-39.19Å. Overall, compound 47 binding limits conformational sampling of HCV RdRp, shifting the enzyme to a more stabilized closed conformation. Compound 47 allowed the thumb and finger subdomains to come into proximity to one another thereby decreasing the width of the template channel, subsequently altering the interdomain angle. In doing so, this creates a difficult task for the RNA template to gain access to the template channel and the active binding site<sup>66</sup>. Based on these findings, it is believed that compound 47 prevents the transitions between closed and open conformations, which

is imperative for RdRp functioning in replication, leading to inadequate and inefficient execution of RNA elongation and subsequent viral replication <sup>7,67</sup>.



**Figure 6:** Distance analysis demonstrating the conformational changes of the TCW and interdomain angle upon compound 47 binding. **(A)** Interdomain angle was computed by measuring the angle between the centre of masses of the finger, thumb and palm domains. **(B)** The width is calculated by measuring the distance between the C- $\alpha$  atoms of residues Met139 and Val405. **(C)** Structural dynamic movements of the TCW in the apo (purple) and bound (orange, ligand highlighted in green) states of HCV RdRp throughout a 200ns molecular dynamics simulation. The dashed line in the highlighted images represents the distance between the C- $\alpha$  atoms of the two residues.

To determine the occurrence of correlated dynamics, DCCM analysis was performed. This method computes the position of the C- $\alpha$  atoms throughout each independent 200ns MD simulation<sup>68</sup>. The colours red to yellow designates highly positive correlated C- $\alpha$  movements. Contrastingly, negative or anti-correlated movements are represented as colours blue to black. While similar correlated motions occur in both cases, the apo enzyme shows higher levels of correlation (red region), further justified by observing a higher occupancy of positive cross-correlation coefficient, as depicted in Figure 7A. This correlation, decreases upon ligand binding (Figure 7B), suggesting selective covalent inhibitor, compound 47, induces a reduction in fluctuations of RdRp C- $\alpha$  atoms. Residues 246-282 and 380-550 which correspond to the finger and thumb domains, respectively, displayed anti-correlated movement (dark blue region) in the apo and to a slightly lesser degree in the bound system, supporting the residue fluctuations in Figure 5. These domain motions serve fundamental importance as it facilitates RdRp closed “active” conformation, which allow for *de novo* synthesis of RNA. Therefore reduction when compound 47 is bound would be consistent with the inability of the enzyme to efficiently execute viral replication as the RdRp will not be adequately equipped to accept and bind RNA template thus hindering the process of viral replication<sup>7</sup>.

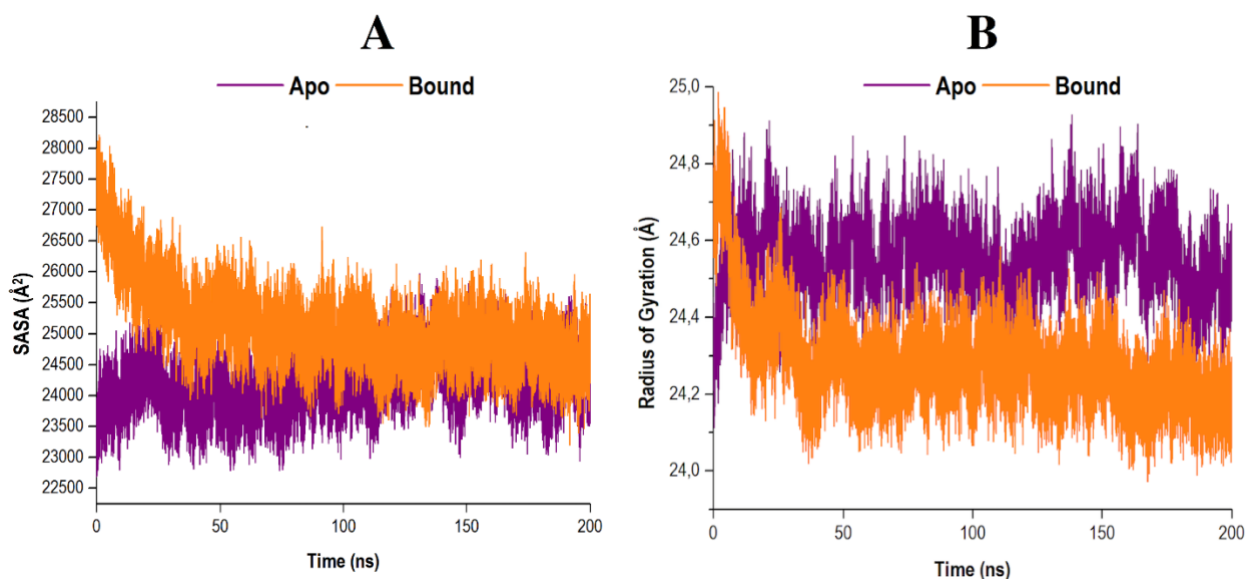


**Figure 7:** Dynamical cross-correlation matrix presenting correlation of residues in the apo (**A**) and compound 47-bound system (**B**). The status of correlated motions is deduced by the colour scale on the right, the black cross indicates Cys366 to which compound 47 is covalently bound.

### 3.3 Solvent Accessible Surface Area and Radius of Gyration

To further clarify the impact of compound 47 binding on the structure of RdRp, the solvent accessible surface area (SASA) was evaluated. Solvent accessibility quantifies the area of a protein obtainable by

or exposed to solvent molecules <sup>69,70</sup>. SASA analysis of both systems were conducted and the results obtained were presented in Figure 8 (A).

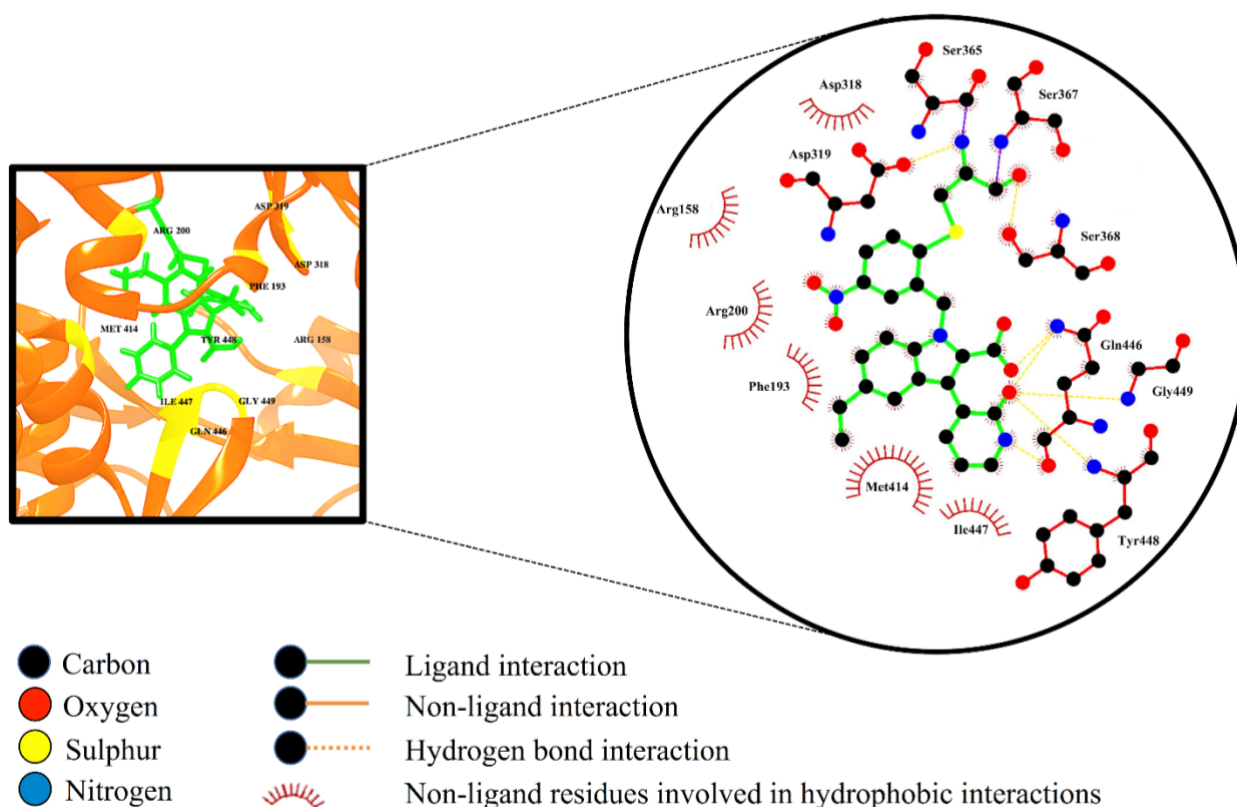


**Figure 8:** (A) Solvent accessible surface area of apo and compound 47 bound RdRp. (B) Radius of gyration of both systems were measured over a 200ns simulation. Data obtained was represented as a plot displaying the differences arising in radius deviation between the apo and covalent systems.

The results indicate that the apo system maintained a consistent residual exposure to solvents throughout the 200ns MD simulation. Contrastingly, the bound system exhibited a progressive decline in the area accessible to solvents upon ligand binding. This may serve as an indication that following compound 47 binding to the RdRp, structural integrity of the protein was compromised. The active site residues underwent restructuring, in the process diminishing the area of which was exposed to solvents. Consequently, these actions may result in the functional loss of proper enzyme activity <sup>71</sup>. The radius of gyration (RoG) measures the compactness of a protein whilst concurrently providing insight into the biological system's stability <sup>46,72,73</sup>. The comparative RoG for the apo and bound systems is represented in Figure 8 (B). From the relevant figure, it is observed that the RoG value for both the apo and bound systems are fairly similar, with an average value of 24.55Å and 24.28Å, respectively. These findings suggest that compound 47 binding to RdRp Cys366 induced a conformational shift to a more compact structure, despite only a 0.27Å change, ultimately forming a relatively stable enzyme.

### 3.4 Ligand interaction with HCV NS5B RdRp

Selective covalent inhibition disrupts atomic backbone dynamics of the RdRp through trepidation of regulatory elements such as the thumb domain,  $\beta$ -loop and the C-terminal tail. Normal physiological structure-based dynamics of the enzyme requires these imperative modulatory components for efficient viral transcription and replication. As observed in Figure 9, the indole core (C<sub>8</sub>H<sub>7</sub>N) of the inhibitor as well as the C-3 pyridone ring are in close contact with Met414. Generally, the active site adopts a favourable morphology to allow influential elements such as the  $\beta$ -loop, a component vital for de novo initiation, to swing away from the active site thereby sanctioning elongation of the RNA product. Upon compound 47 binding, the C-3 pyridone ring of the inhibitor forms three hydrogen interactions with the backbone of  $\beta$ -loop residues Gln446, Tyr448 and Gly449. These bond interactions induce the descent of the  $\beta$ -loop further into the palm site thereby disrupting the structural conformation of RdRp, giving rise to loss of structural integrity. Ultimately, these consequences may result in the overall impediment of RdRp enzymatic activity <sup>74</sup>.



**Figure 9:** Ligand-residue interaction diagram of compound, 47 inside the active site of the palm domain of the RdRp. Orange dashed lines denote hydrogen bond interactions.

### 3.5 Free Binding Energy Calculations

The Molecular Mechanics/Generalized-Born Surface Area (MM/GBSA) method is extensively utilised to approximate the BFE of an inhibitor at an atomic level <sup>75</sup>. The MM/GBSA method was used to evaluate the total binding energy contributions of compound 47 to the RdRp as presented in Table 1.

**Table 1:** Summary of MM/GBSA-based binding free energy contributions to the compound 47- HCV RdRp complex

Energy components (kcal.mol <sup>-1</sup> )					
	$\Delta E_{VDW}$	$\Delta E_{elec}$	$\Delta G_{gas}$	$\Delta G_{solv}$	$\Delta G_{bind}$
<b>HCV RdRp</b>	-4537.81 ± 26.86	-35 742.39 ± 139.77	-4144.35 ± 146.43	-7040.47 ± 115.57	-11 184.82 ± 79.80
<b>Compound 47</b>	-7.30 ± 2.0	-9.39 ± 2.45	78.08 ± 6.48	-37.53 ± 1.20	40.55 ± 6.36
<b>Complex</b>	-47.77 ± 2.92	-160.62 ± 7.68	-129.92 ± 8.14	60.94 ± 5.92	-68.98 ± 4.55

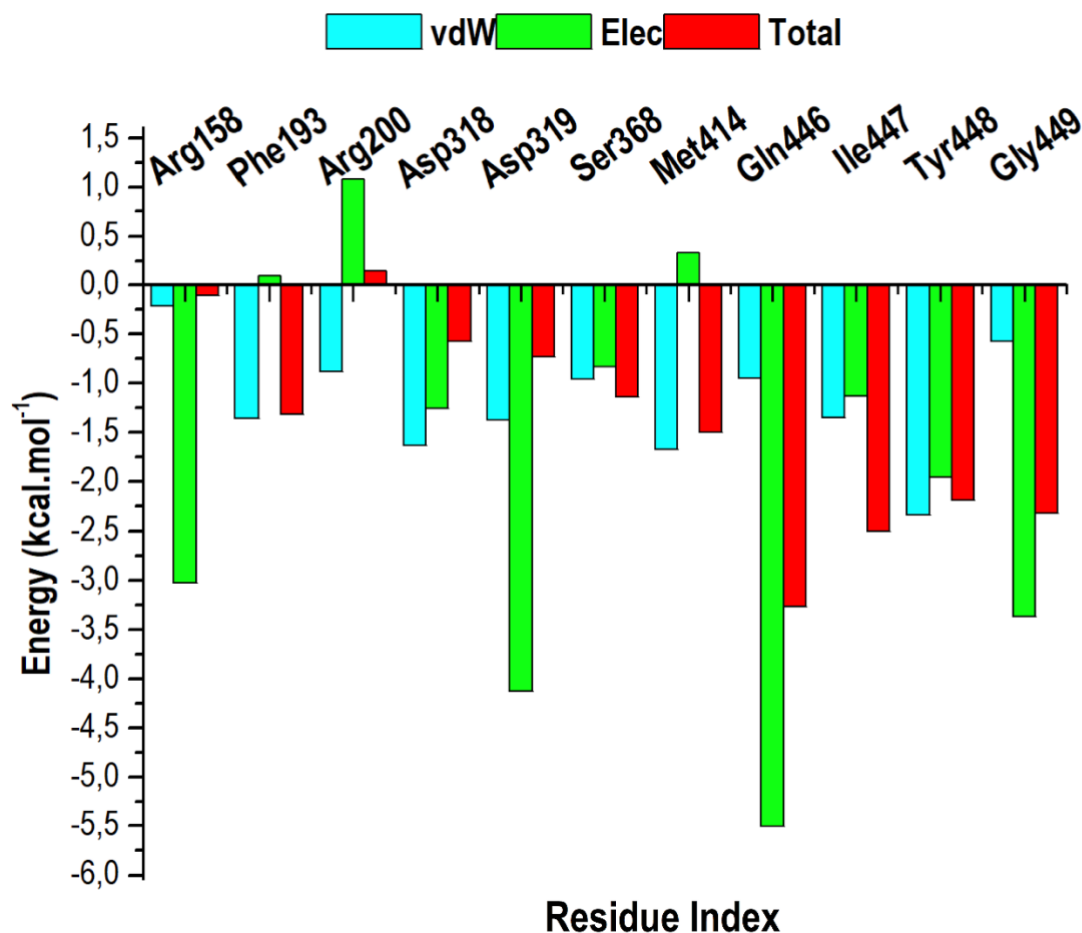
$\Delta E_{VDW}$ : van der Waal;  $\Delta E_{elec}$ : electrostatic;  $\Delta G_{gas}$ : gas phase interaction;  $\Delta G_{solv}$ : solvation free energy;  $\Delta G_{bind}$ : calculated total free binding energy

The calculated free binding energies provide inclusive evidence at a molecular perspective. This technique delivers beneficial data to lay groundwork that can propel the design and discovery of small inhibitory molecules that possess enhanced ligand binding properties. The calculations are dynamic, inexpensive and can be conducted by members of the scientific community as the process is programmed using an external interface server and required software is easily accessible online. The calculated free binding energy of the compound 47 and RdRp system was -68.98 kcal.mol<sup>-1</sup>. Interaction forces such as van der Waals (-47.77 ± 2.92 kcal.mol<sup>-1</sup>) and electrostatic (-160.62 ± 7.68 kcal.mol<sup>-1</sup>) vastly contribute toward the total binding energy of compound 47 to RdRp. Hydrophobic residues lining the binding site pocket contributes significantly to the covalently bound systems free binding energy.



### 3.6 Per-residue Interaction Energy Decomposition Analysis

The total BFE for compound 47 was decomposed into individual residue-based contributions using the MM/GBSA approach. The van der Waals (vdW) and electrostatic (elec) interaction contributions relative to BFE of compound 47 to RdRp was estimated to acquire perspective into which residue and energy constituents had an overall greater impact on the total binding energy (Total). Per-residue energy decomposition analysis was executed and presented as Figure 10.



**Figure 10:** The per-residue energy decomposition analysis of compound 47 bound HCV RdRp.

Intermolecular interactions between the residues of the active site facilitate the binding and stabilisation of compound 47 in the hydrophobic pocket of RdRp. As observed in figure 10, residues that contributes the most energy toward the complex include Arg158 (-3.02 kcal.mol<sup>-1</sup> [elec]), Asp319 (-4.125 kcal.mol<sup>-1</sup> [elec]), Gln446 (-5.51 kcal.mol<sup>-1</sup> [elec]), Tyr448 (-2.35 kcal.mol<sup>-1</sup> [vdW]; -1.96 kcal.mol<sup>-1</sup> [elec]) and Gly449 (-3.37 kcal.mol<sup>-1</sup> [elec]). While the binding site residues that contributed less energy towards the complex were: Phe193 (-1.35 kcal.mol<sup>-1</sup> [vdw]; 0.10 kcal.mol<sup>-1</sup> [elec]), Arg200 (-1.08 kcal.mol<sup>-1</sup>

[vdW]; 1.08 kcal.mol<sup>-1</sup> [elec]), Asp318 (-1.64 kcal.mol<sup>-1</sup> [vdW]; -1.23 kcal.mol<sup>-1</sup> [elec]), Ser368 (-0.96 kcal.mol<sup>-1</sup> [vdW]; -0.83 kcal.mol<sup>-1</sup> [elec]), and Met414 (-1.68 kcal.mol<sup>-1</sup> [vdW]; -0.33 kcal.mol<sup>-1</sup> [elec]). From Figure 10, it can be deduced that electrostatic interactions from residues 158, 319, 446, 449 and van der Waals interactions from 448 and 414 attributed to the high-energy interaction of the covalent system ( $\Delta G_{\text{Bind}}$ : -68.98 kcal.mol<sup>-1</sup>).

#### **4 Conclusion**

The HCV RdRp has a pivotal role as the driving force behind viral transcription and replication. Current market-available drugs exhibit rigid pharmacodynamic profiles that require optimization for efficient viral inhibition in a safe yet effective manner. The conceptualization of selective covalent inhibition brought a new-targeted approach to HCV antiviral therapeutics. Compound 47, an indole core small inhibitor, binds covalently to Cys366 of the RdRp. The consequential shift in conformation prompted by selective covalent binding was led by modifications in enzyme flexibility, correlated dynamics and intermolecular bonding. Taken together, dynamic and structural effects of compound 47 binding disseminate throughout the RdRp, affecting structural integrity of the enzyme. This may lead to the loss of appropriate RdRp functioning, ultimately preventing viral replication and propagation. Therefore, it can be concluded that selective covalent inhibition is a promising approach that has the potential to facilitate future potent anti-HCV therapeutic strategies.

#### **Acknowledgements**

Authors acknowledge the German Academic Exchange Service (DAAD) and the National Research Foundation (NRF) and The Centre of High Performance Computing (CHPC, [www.chpc.ac.za](http://www.chpc.ac.za)), Cape Town, RSA, for computational resources.

#### **Conflict of Interest**

Authors declare no potential conflicts of interest.

## References

- 1 T. Kazakov, F. Yang, H. N. Ramanathan, A. Kohlway, M. S. Diamond and B. D. Lindenbach, *PLoS Pathog.*, 2015, 11, e1004817.
- 2 Q. L. Choo, G. Kuo, A. J. Weiner, L. R. Overby, D. W. Bradley and M. Houghton, *Science* (80-), 1989, 244, 359–362.
- 3 M. J. Karoney, A. M. Siika and M. J. Karoney, *Pan African Med. Journal. Pan African Med. J.*, 2013, 141444, 44–1937.
- 4 K. H. Barakat, J. Law, A. Prunotto, W. C. Magee, D. H. Evans, D. L. Tyrrell, J. Tuszynski and M. Houghton, *J. Chem. Inf. Model.*, 2013, 53, 3031–3043.
- 5 S. Sun, V. B. Rao and M. G. Rossmann, *Curr. Opin. Struct. Biol.*, 2010, 20, 114–120.
- 6 K. X. Chen, C. A. Lesburg, B. Vibulbhan, W. Yang, T. Y. Chan, S. Venkatraman, F. Velazquez, Q. Zeng, F. Bennett, G. N. Anilkumar, J. Duca, Y. Jiang, P. Pinto, L. Wang, Y. Huang, O. Selyutin, S. Gavalas, H. Pu, S. Agrawal, B. Feld, H. C. Huang, C. Li, K. C. Cheng, N. Y. Shih, J. A. Kozlowski, S. B. Rosenblum and F. G. Njoroge, *J. Med. Chem.*, 2012, 55, 2089–2101.
- 7 A. A. Eltahla, E. Tay, M. W. Douglas and P. A. White, *Antimicrob Agents Chemother*, 2014, 58, 7215–7224.
- 8 E. Sesmero and I. F. Thorpe, *Viruses*, 2015, 7, 3974–3994.
- 9 Y. Wei, J. Li, J. Qing, M. Huang, M. Wu, F. Gao, D. Li, Z. Hong, L. Kong, W. Huang and J. Lin, *PLoS One*, 2016, 11, e0148181.
- 10 M. L. Barreca, N. Iraci, G. Manfroni and V. Cecchetti, *Future Med. Chem.*, 2011, 3, 1027–1055.
- 11 K. X. Chen, B. Vibulbhan, W. Yang, M. Sannigrahi, F. Velazquez, T. Y. Chan, S. Venkatraman, G. N. Anilkumar, Q. Zeng, F. Bennet, Y. Jiang, C. A. Lesburg, J. Duca, P. Pinto, S. Gavalas, Y. Huang, W. Wu, O. Selyutin, S. Agrawal, B. Feld, H. C. Huang, C. Li, K. C. Cheng, N. Y. Shih, J. A. Kozlowski, S. B. Rosenblum and F. G. Njoroge, *J. Med. Chem.*, 2012, 55, 754–765.
- 12 C. B. Dousson, *Antivir. Chem. Chemother.*, 2018, 26, 204020661875643.
- 13 E. Lawitz, M. Buti, J. M. Vierling, P. L. Almasio, S. Bruno, P. J. Ruane, T. I. Hassanein, B. Muellhaupt, B. Pearlman, L. Jancoriene, W. Gao, H. C. Huang, A. Shepherd, B. Tannenbaum, D. Fernsler, J. J. Li, A. Grandhi, H. Liu, F. H. Su, S. Wan, F. J. Dutko, B. Y. T. Nguyen, J.

- Wahl, M. N. Robertson, E. Barr, W. W. Yeh, R. M. Plank, J. R. Butterson and E. M. Yoshida, *Lancet Gastroenterol. Hepatol.*, 2017, 2, 814–823.
- 14 D. Wyles, H. Wedemeyer, Z. Ben-Ari, E. J. Gane, J. B. Hansen, I. M. Jacobson, A. L. Laursen, A. Luetkemeyer, R. Nahass, S. Pianko, S. Zeuzem, P. Jumes, H. C. Huang, J. Butterson, M. Robertson, J. Wahl, E. Barr, H. K. Joeng, E. Martin and L. Serfaty, *Hepatology*, 2017, 66, 1794–1804.
- 15 B. Devogelaere, J. M. Berke, L. Vijgen, P. Dehertogh, E. Fransen, E. Cleiren, L. Van Der Helm, O. Nyanguile, A. Tahri, K. Amssoms, O. Lenz, M. D. Cummings, R. F. Clayton, S. Vendeville, P. Rabisson, K. A. Simmen, G. C. Fanning and T. I. Lin, *Antimicrob. Agents Chemother.*, 2012, 56, 4676–4684.
- 16 A. A. Eltahla, F. Luciani, P. A. White, A. R. Lloyd and R. A. Bull, *Viruses*, 2015, 7, 5206–5224.
- 17 J. Li, X. Liu, S. Li, Y. Wang, N. Zhou, C. Luo, X. Luo, M. Zheng, H. Jiang and K. Chen, *Int. J. Mol. Sci.*, 2013, 14, 22845–22856.
- 18 I. Buxton and L. Benet, *Goodman Gilman’s Pharmacol. Basis Ther.*, 2011, 17–39.
- 19 K. Kumthip and N. Maneekarn, *Virolog. J.*, 2015, 12, 1–12.
- 20 N. Echeverria, G. Moratorio, J. Cristina and P. Moreno, *World J. Hepatol.*, 2015, 7, 831–845.
- 21 S. Fonseca-Coronado, A. Escobar-Gutiérrez, K. Ruiz-Tovar, M. Y. Cruz-Rivera, P. Rivera-Osorio, M. Vazquez-Pichardo, J. C. Carpio-Pedroza, J. A. Ruíz-Pacheco, F. Cazares and G. Vaughan, *J. Clin. Microbiol.*, 2012, 50, 281–287.
- 22 D. S. Johnson, E. Weerapana and B. F. Cravatt, *Future Med. Chem.*, 2010, 2, 949–964.
- 23 R. A. Bauer, *Drug Discov. Today*, 2015, 20, 1061–1073.
- 24 G. Lee, D. E. Piper, Z. Wang, J. Anzola, J. Powers, N. Walker and Y. Li, *J. Mol. Biol.*, 2006, 357, 1051–1057.
- 25 M. Hagel, D. Niu, T. St Martin, M. P. Sheets, L. Qiao, H. Bernard, R. M. Karp, Z. Zhu, M. T. Labenski, P. Chaturvedi, M. Nacht, W. F. Westlin, R. C. Petter and J. Singh, *Nat. Chem. Biol.*, 2011, 7, 22–24.
- 26 K. K. Hallenbeck, D. M. Turner, A. R. Renslo and M. R. Arkin, *Curr. Top. Med. Chem.*, 2017, 17, 4–15.
- 27 H. McWilliam, W. Li, M. Uludag, S. Squizzato, Y. M. Park, N. Buso, A. P. Cowley and R. Lopez, *Nucleic Acids Res.*, 2013, 41, W597–W600.

- 28 A. Bateman, M. J. Martin, C. O'Donovan, M. Magrane, E. Alpi, R. Antunes, B. Bely, M. Bingley, C. Bonilla, R. Britto, B. Bursteinas, H. Bye-AJee, A. Cowley, A. Da Silva, M. De Giorgi, T. Dogan, F. Fazzini, L. G. Castro, L. Figueira, P. Garmiri, G. Georghiou, D. Gonzalez, E. Hatton-Ellis, W. Li, W. Liu, R. Lopez, J. Luo, Y. Lussi, A. MacDougall, A. Nightingale, B. Palka, K. Pichler, D. Poggioli, S. Pundir, L. Pureza, G. Qi, S. Rosanoff, R. Saidi, T. Sawford, A. Shypitsyna, E. Speretta, E. Turner, N. Tyagi, V. Volynkin, T. Wardell, K. Warner, X. Watkins, R. Zaru, H. Zellner, I. Xenarios, L. Bougueleret, A. Bridge, S. Poux, N. Redaschi, L. Aimo, G. ArgoudPuy, A. Auchincloss, K. Axelsen, P. Bansal, D. Baratin, M. C. Blatter, B. Boeckmann, J. Bolleman, E. Boutet, L. Breuza, C. Casal-Casas, E. De Castro, E. Coudert, B. Cuche, M. Doche, D. Dornevil, S. Duvaud, A. Estreicher, L. Famiglietti, M. Feuermann, E. Gasteiger, S. Gehant, V. Gerritsen, A. Gos, N. Gruaz-Gumowski, U. Hinz, C. Hulo, F. Jungo, G. Keller, V. Lara, P. Lemercier, D. Lieberherr, T. Lombardot, X. Martin, P. Masson, A. Morgat, T. Neto, N. Nospikel, S. Paesano, I. Pedruzzi, S. Pilbout, M. Pozzato, M. Pruess, C. Rivoire, B. Roechert, M. Schneider, C. Sigrist, K. Sonesson, S. Staehli, A. Stutz, S. Sundaram, M. Tognolli, L. Verbregue, A. L. Veuthey, C. H. Wu, C. N. Arighi, L. Arminski, C. Chen, Y. Chen, J. S. Garavelli, H. Huang, K. Laiho, P. McGarvey, D. A. Natale, K. Ross, C. R. Vinayaka, Q. Wang, Y. Wang, L. S. Yeh and J. Zhang, *Nucleic Acids Res.*, 2017, 45, D158–D169.
- 29 J. de Vicente, R. T. Hendricks, D. B. Smith, J. B. Fell, J. Fischer, S. R. Spencer, P. J. Stengel, P. Mohr, J. E. Robinson, J. F. Blake, R. K. Hilgenkamp, C. Yee, G. Adjabeng, T. R. Elworthy, J. Li, B. Wang, J. T. Bamberg, S. F. Harris, A. Wong, V. J. P. Leveque, I. Najera, S. Le Pogam, S. Rajyaguru, G. Ao-Ieong, L. Alexandrova, S. Larrabee, M. Brandl, A. Briggs, S. Sukhtankar and R. Farrell, *Bioorganic Med. Chem. Lett.*, 2009, 19, 5652–5656.
- 30 J. Bhachoo and T. Beuming, in *Methods in Molecular Biology*, 2017, vol. 1561, pp. 235–254.
- 31 G. Madhavi Sastry, M. Adzhigirey, T. Day, R. Annabhimoju and W. Sherman, *J. Comput. Aided. Mol. Des.*, 2013, 27, 221–234.
- 32 Z. Yang, K. Lasker, D. Schneidman-Duhovny, B. Webb, C. C. Huang, E. F. Pettersen, T. D. Goddard, E. C. Meng, A. Sali and T. E. Ferrin, *J. Struct. Biol.*, 2012, 179, 269–278.
- 33 O. Trott and A. J. Olson, *J. Comput. Chem.*, 2010, 31, 455–61.
- 34 E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, *J. Comput. Chem.*, 2004, 25, 1605–1612.
- 35 F. Fogolari, A. Corazza, S. Toppo, S. C. E. Tosatto, P. Viglino, F. Ursini and G. Esposito, *J. Biomed. Biotechnol.*, 2012, 2012, 303190.

- 36 D. a. Case, I. T.E. Cheatham, T. Darden, H. Gohlke, R. Luo, J. K.M. Merz, a. Onufriev, C. Simmerling, B. Wang and R. Woods., *J. Comput. Chem*, 2005, 26, 1668–1688.
- 37 R. Salomon-Ferrer, D. A. Case and R. C. Walker, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2013, 3, 198–210.
- 38 P. Ramharack, S. Oguntade and M. E. S. Soliman, *RSC Adv.*, 2017, 7, 22133–22144.
- 39 S. Khan, I. Bjjj, R. M. Betz and M. E. S. Soliman, *Future Med. Chem.*, 2018, 10, 1003–1015.
- 40 S. Khan, I. Bjjj, F. Oluto and M. E. S. Soliman, *Future Med. Chem.*, 2018, 10, 2265–2275.
- 41 R. Betz, Dabble, <https://dabble.robinbetz.com/>, 2017
- 42 N. N. Mhlongo, M. Ebrahim, A. A. Skelton, H. G. Kruger, I. H. Williams and M. E. S. Soliman, *RSC Adv.*, 2015, 5, 82381–82394.
- 43 P. Gonnet, *J. Comput. Phys.*, 2007, 220, 740–750.
- 44 W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graph.*, 1996, 14, 33–38.
- 45 A. Bornot, C. Etchebest and A. G. De Brevern, *Proteins Struct. Funct. Bioinforma.*, 2011, 79, 839–852.
- 46 M. I. Lobanov, N. S. Bogatyreva and O. V Galzitskaia, *Mol. Biol. (Mosk.)*, 2008, 42, 701–706.
- 47 V. Gapsys, S. Michielssens, J. H. Peters, B. L. de Groot and H. Leonov, *Methods Mol. Biol.*, 2015, 1215, 173–209.
- 48 M. Schauerl, P. Czodrowski, J. E. Fuchs, R. G. Huber, B. J. Waldner, M. Podewitz, C. Kramer and K. R. Liedl, *J. Chem. Inf. Model.*, 2017, 57, 345–354.
- 49 C. Agoni, P. Ramharack and M. E. S. Soliman, *Future Med. Chem.*, 2018, 10, 1665–1675.
- 50 J. M. Hayes and G. Archontis, in *Molecular Dynamics – Studies of Synthetic and Biological Macromolecules*, 2011.
- 51 T. Hou, J. Wang, Y. Li and W. Wang, *J. Chem. Inf. Model.*, 2011, 51, 69–82.
- 52 X. Zhang, H. Perez-Sanchez and F. C. Lightstone, *Curr. Top. Med. Chem.*, 2017, 17, 1631–1639.
- 53 B. R. Miller, T. D. McGee, J. M. Swails, N. Homeyer, H. Gohlke and A. E. Roitberg, *J. Chem. Theory Comput.*, 2012, 8, 3314–3321.
- 54 N. Singh and A. Warshel, *Proteins Struct. Funct. Bioinforma.*, 2010, 78, 1705–1723.
- 55 P. Ramharack and M. E. S. Soliman, *J. Biomol. Struct. Dyn.*, 2018, 36, 1118–1133.

- 56 K. E. Machaba, N. N. Mhlongo, Y. M. Dokurugu and M. E. S. Soliman, *Future Med. Chem.*, 2017, 9, 1055–1071.
- 57 U. Ndagi, N. N. Mhlongo and M. E. Soliman, *Mol. BioSyst.*, 2017, 13, 1157–1171.
- 58 S. Oguntade, P. Ramharack and M. E. Soliman, *Future Virol.*, 2017, 12, 261–273.
- 59 P. Appiah-Kubi and M. E. S. Soliman, *J. Biomol. Struct. Dyn.*, 2016, 34, 2418–2433.
- 60 R. Galindo-Murillo, D. R. Roe and T. E. Cheatham, *Biochim. Biophys. Acta - Gen. Subj.*, 2015, 1850, 1041–1058.
- 61 B. Hess, *Phys. Rev. E - Stat. Physics, Plasmas, Fluids, Relat. Interdiscip. Top.*, 2002, 65, 031910.
- 62 C. V. Kumar, R. G. Swetha, A. Anbarasu and S. Ramaiah, *Adv. Bioinformatics*, 2014, 2014, 1–10.
- 63 K. Henzler-Wildman and D. Kern, *Nature*, 2007, 450, 964–972.
- 64 L. Q. Yang, P. Sang, Y. Tao, Y. X. Fu, K. Q. Zhang, Y. H. Xie and S. Q. Liu, *J. Biomol. Struct. Dyn.*, 2014, 32, 372–393.
- 65 B. C. Davis, J. A. Brown and I. F. Thorpe, *Biophys. J.*, 2015, 108, 1785–1795.
- 66 B. C. Davis, J. A. Brown and I. F. Thorpe, *Biophys. J.*, 2015, 108, 1785–1795.
- 67 J. A. Brown and I. F. Thorpe, *Biochemistry*, 2015, 54, 4131–4141.
- 68 U. Ndagi, N. N. Mhlongo and M. E. Soliman, *Mol. BioSyst.*, 2017, 13, 1157–1171.
- 69 T. J. Richmond, *J. Mol. Biol.*, 1984, 178, 63–89.
- 70 L. McGillewie and M. E. Soliman, *Mol. Biosyst.*, 2016, 12, 1457–1467.
- 71 S. E. Boyce, N. Tirunagari, A. Niedziela-Majka, J. Perry, M. Wong, E. Kan, L. Lagpacan, O. Barauskas, M. Hung, M. Fenaux, T. Appleby, W. J. Watkins, U. Schmitz and R. Sakowicz, *PLoS One*, 2014, 9, e84808.
- 72 T. Sindhu and P. Srinivasan, *RSC Adv.*, 2015, 5, 14202–14213.
- 73 L. McGillewie and M. E. Soliman, *Mol. Biosyst.*, 2016, 12, 1457–1467.
- 74 J. Li and K. A. Johnson, *J. Biol. Chem.*, 2016, 291, 10067–10077.
- 75 S. Genheden and U. Ryde, *Expert Opin. Drug Discov.*, 2015, 10, 449–461.

## CHAPTER 6

### **Enhanced Pharmacophore-based Virtual Screening Approach in the Discovery of Potential Inhibitors of HCV NS3 Helicase**

Letitia Shunmugam<sup>a</sup>, Nikita Devnarain<sup>a</sup> and Mahmoud E. S. Soliman<sup>a\*</sup>

<sup>a</sup>Molecular Bio-Computation and Drug Design laboratory, School of Health Sciences, University of KwaZulu-Natal, Westville Campus, Durban 4001, South Africa.

\*Corresponding Author: Mahmoud E.S. Soliman

School of Health Sciences, University of KwaZulu-Natal, Westville Campus, Durban 4001, South Africa

Email: soliman@ukzn.ac.za

Telephone: +27 (0) 31 260 7413, Fax: +27 (0) 31 260 7872



## **Abstract**

The helicase is essential in the modulation of Hepatitis C virus (HCV) life cycle, hence a good target for drug design, a domain highly neglected in literature. Quercetin is an inhibitor of several viruses, including *flaviviruses*. However due to its compromised pharmacodynamics, development of new potent helicase inhibitors that address challenges in antiviral therapies are crucial. Herein, an in-house pharmacophore-based virtual screening protocol was applied to identify improved potential inhibitors based on highly contributing residues that bind with highest affinity to quercetin. This is a preliminary investigation to initiate wider-spectrum screening of potential helicase inhibitors and can be implemented in rational drug design of anti-HCV drugs and other viruses. Further computational and experimental work are advised to validate this approach.

## **Keywords**

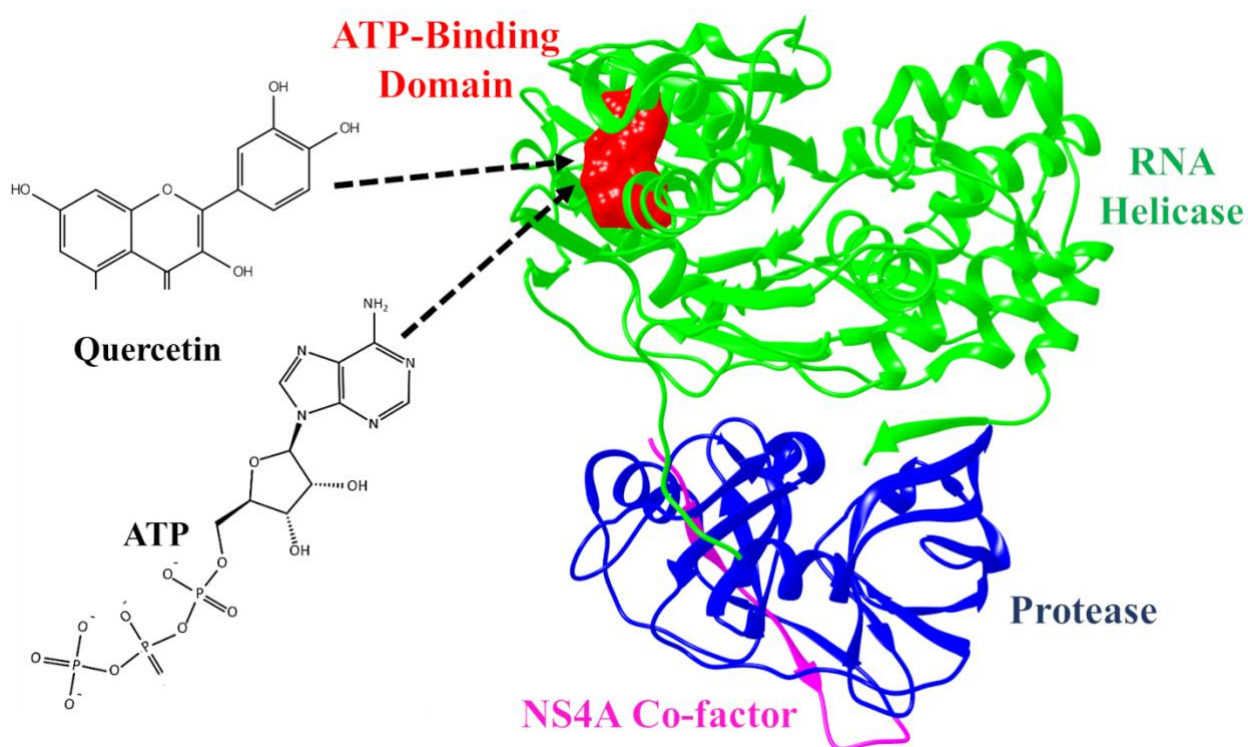
Hepatitis C virus, drug discovery; NS3 helicase potential inhibitors; per-residue energy decomposition-based pharmacophore; virtual screening.

## 1. Introduction

Hepatitis C virus (HCV) is the causative agent of hepatitis C infection [1]. Identified as a hepatropic single-stranded RNA virus, HCV is a member of the *Flaviviridae* family from the *Hepacivirus* genus [2]. The virus leads to major hepatic-related disorders such as cirrhosis, fibrosis and hepatocellular carcinomas, resulting in severe liver damage and in some cases, death [3]. The World Health Organization (WHO) has estimated that approximately 170-185 million people are affected by HCV throughout the world and is fast becoming a global pandemic [4].

Drug development industries have focused intensively on viral enzymes as possible therapeutic targets based on their important roles in the viral life cycle [5]. One of the most significant HCV enzymes is the non-structural protein 3 (NS3), which consists of the RNA helicase, non-structural protein 4A (NS4A) co-factor and serine protease [6–8]. At present, few NS3 protease inhibitors have been approved by the Food and Drug Administration (FDA), such as telaprevir [9], boceprevir [10], simeprevir [11], elbasvir/grazoprevir [12] and danoprevir [13]. Although the NS3 protease serves as a prevalent target in antiviral therapy, drug resistance and system toxicity has created an ever-escalating obstacle in drug development, consequently igniting the need to explore and discover novel inhibitors against HCV [14].

The NS3 helicase is often a neglected aspect of the NS3 enzyme complex as most attention is directed toward the protease. In the HCV life cycle, the helicase plays a vital role in the execution of genomic replication and appropriate viral assembly [8]. Helicases are described as small molecular motors driven by the hydrolysis of adenosine triphosphate (ATP), a process that allows the helicase to grasp a single RNA or DNA strand, causing dissociation from its complementary strand. The RNA helicase is an essential requirement for cell transcription, translation, and RNA splicing [15], thus rendering the protein an attractive candidate for drug target studies [8,16].



**Figure 1:** Three-dimensional structure of NS3 protein complex (PDB: 3O8D). The ATP-binding domain (highlight in red) is the region in which ATP or Quercetin can effectively bind.

Quercetin is a natural flavonoid compound found in most vegetables and fruits [17]. The flavonoid has been identified as a competitive inhibitor of adenosine triphosphate (ATP) at the relative binding site located within the viral RNA helicase (Figure 1). This phenomena was successfully observed in *flavivirus* such as Dengue fever as well as HCV [16,18]. The therapeutic utilization of quercetin faces major difficulties; the compound is relatively unstable within a physiological intermediate and has poor solubility, thus limiting its application solely to oral administration [19]. For these reasons, it would be highly beneficial to discover potential compounds that can inhibit ATP binding with similar moieties, improved pharmacodynamic profiles, higher potency and restricted adverse effects.

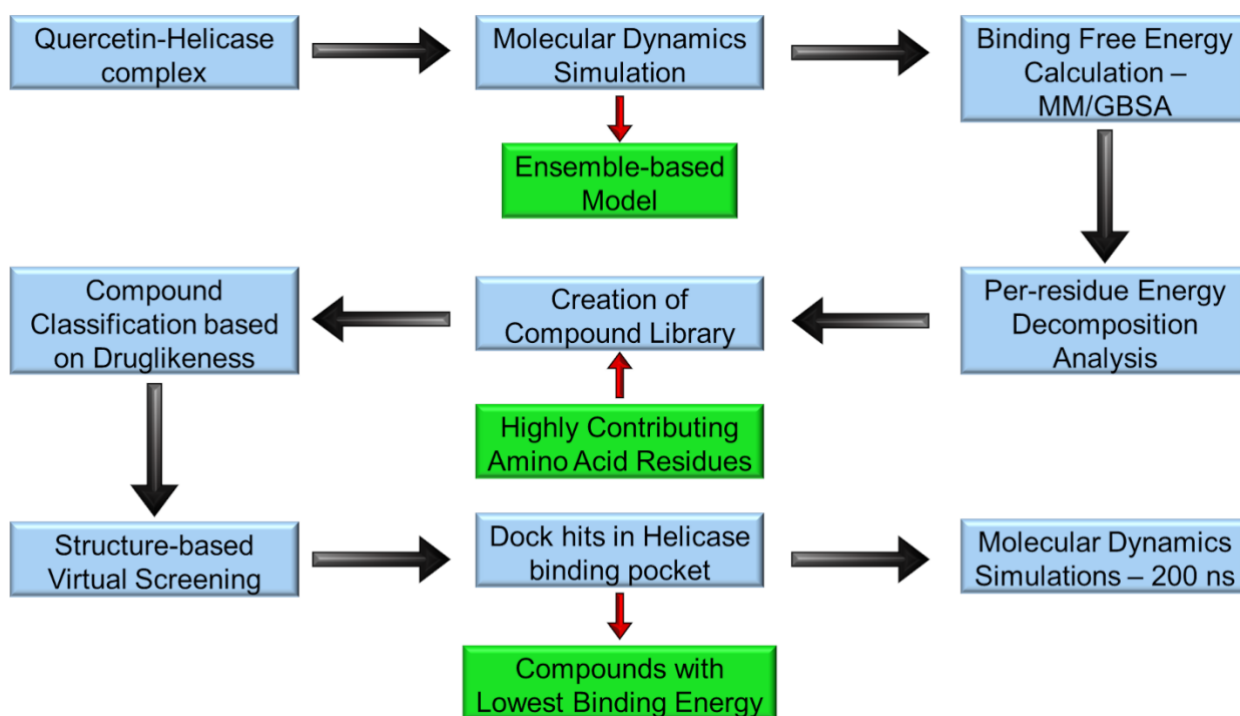
Over the years, computer-aided drug design (CADD) methods have greatly contributed toward the design and development of drugs against a wide diversity of ailments, viruses and bacteria [20]. Computational practices are often utilized to accelerate the discovery process and direct attention toward compounds that exhibit potential antiviral activities, elaborate on enzyme-drug interactions and provide improvements for promising compounds [21,22].

Virtual screening (VS) is a process that uses a sequence of computational tools to search through compound libraries or databases for small molecules with similar structures to the template scaffold.

The structures found through VS should bind to the amino acid residues within the active site of the target [23–26]. Structure-based VS screens for potential ligands based on the 3D-structure of an enzyme’s active binding pocket. Contrastingly, pharmacophore or ligand-based VS searches for favorable, similar features of ligands with known bioactive properties for pharmacophore model generation [27]. The latter approach is based on intermolecular interactions that include hydrogen bonding, electrostatic energies and charge. This approach also screens for potential compounds that form interactions with highly contributing residues based on the binding free energy (BFE) [28].

Herein, we implemented molecular dynamics (MD) simulations, pharmacophore-based VS and per-residue energy decomposition (PRED) analysis for residues that contribute the most to free binding energy, to enhance the precision of the pharmacophore model. The pharmacophoric approach followed in this study can be implemented by medicinal chemists in the drug discovery process for potent inhibitors against HCV.

## 2. Computational Methodology



**Figure 2.** The systematic workflow applied in this study.

### 2.1 Receptor and Ligand Preparation

The HCV NS3 helicase-protease complex was obtained from RSCB Protein Data Bank (PDB code: 3O8D) [29]. The 3D-structure of quercetin was created using *Avogadro* [30]. The helicase and ligands were prepared using UCSF Chimera software [31].

## ***2.2 Molecular Docking***

The compounds were docked into the ATPase pocket of the HCV NS3 helicase. Through UCSF Chimera, using the AutoDock Vina plugin, compounds were docked by means of the grid box function to express the binding site (spacing of 0.375 Å; x, y, z dimensions of 14.28 x 22.51 x 18.29) [32,33]. The docked results with optimum poses and highest binding affinities to the helicase ATPase pocket were selected and thereafter subjected to MD simulations.

## ***2.3 Molecular Dynamics Simulation***

The utilization of MD simulations provides a powerful tool for investigation of the physical activities executed by atoms and molecules, thus providing an insight into the biomolecular processes of target systems [34]. The MD simulations were performed using the PMEMD dynamics engine in the AMBER14 package with GPU acceleration [35]. The NS3 enzyme was dehydrogenated and parameterized using the AMBER FF14SB force field [36]. Hydrogen atoms were added to the ligands and charged by means of Gasteiger charges. The antechamber module was utilized to generate partial atomic charges for the ligands by applying General Amber Force Field (GAFF) and restrained electrostatic potential (RESP) procedures [37]. The LEAP module was applied to neutralise and solvate the systems through the addition of hydrogen, sodium and chloride atoms, whilst suspending them in an orthorhombic box of TIP3P water molecules so that all atoms were within 10 Å of the box edges. The system was initially minimized with the restraint potential of 10 kcal.mol<sup>-1</sup> Å<sup>-2</sup> for consideration of solute molecule of 2 500 steps for 1 000 steepest descent steps, followed by 1000 steps of conjugate gradient minimization. Additionally, an unrestrained full minimization of 200 steps was completed by conjugate gradient algorithm. A canonical ensemble simulation was performed from 0 to 300 Kelvin (K) for 50ps to maintain a fixed volume and number of atoms in each system. Potential harmonic restraint of 10 kcal.mol<sup>-1</sup> Å<sup>2</sup> for solute atoms and Langevin thermostat with collision frequency of 1.0 ps<sup>-1</sup> was applied to each system. All systems were equilibrated for 500 ps with a constant operating temperature of 300 K. The pressure and the number of atoms were kept constant to bear resemblance to an isobaric–isothermal ensemble (NPT). The pressure the systems were maintained at 1 bar using the Berendsen barostat [38] and the SHAKE algorithm was utilized for hydrogen bond constraint [39,40].

## ***2.4 Thermodynamic Binding Free Energy Calculations***

Molecular dynamic simulations allow free energy differences which controls the underlying mechanism of all biological processes [41]. Calculations for BFE is a crucial method employed to comprehensively observe the binding mechanism between a ligand or compound and enzyme, whilst encompassing both enthalpic and entropic contributions [42]. Estimation of the binding affinity within a docked system is calculated through BFE using the Molecular Mechanics/Generalized Born Surface Area (MM/GBSA) method. This method provides reproducible relative binding affinities of compounds with great accuracy and warrants considerably less computational resources in comparison to a full-scale MD free energy perturbation/thermodynamic simulation [43]. In this study, BFE was averaged over 20 000 snapshots, which were generated from the 200ns trajectory. The explicit solvent employed in the MD simulation was removed and substituted with a dielectric continuum as per MM/GBSA protocol [44]. Molecular mechanics force fields were employed to calculate energy contributions from the atomic coordinates of the ligands, NS3 enzyme and complex in a gas-phase. The equations below elaborate the process in which the binding free energies ( $\Delta G$ ) were assessed:

$$(1) \Delta G_{bind} = \Delta G_{complex} - \Delta G_{receptor} - \Delta G_{ligand}$$

$$(2) \Delta G_{bind} = E_{gas} + G_{sol} - T\Delta S$$

$$(3) E_{gas} = E_{int} + E_{vdw} + E_{ele}$$

$$(4) G_{sol} = G_{GB} + G_{SA}$$

$$(5) G_{SA} = \gamma SASA$$

Where:  $\Delta G_{bind}$  is gas-phase summation, gas-phase energy ( $E_{gas}$ ) and the solvation energy,  $G_{sol}$ , is less the entropy ( $T\Delta S$ ) term. The  $E_{gas}$  is the sum of internal energy,  $E_{int}$ , van der Waals (vdW) energy,  $E_{vdw}$  and electrostatic energy,  $E_{ele}$ . The total solvation energy is calculated by a summation of the total energy contributions of polar and non-polar states ( $G_{GB}$  and  $G_{SA}$ , respectively). The  $G_{SA}$  is calculated by means of the solvent accessible surface area, generated by water probe radius of 1.4 Å. Resolution of  $G_{GB}$  equation allows the determination for the energy contributions of polar states. The total entropy of the solute is denoted as ‘S’ and temperature as ‘T’. The solute and solvent dielectric constants are set to 1 and 80, respectively [45,46].

## 2.5 Creation of Pharmacophore Model and Library Generation

Quercetin, the preliminary potential inhibitor was simulated within the ATP-binding pocket for 20ns to establish its bound conformation (binding affinity: -8.1 kcal.mol<sup>-1</sup>). To create a pharmacophore model, PRED analysis was used to identify the amino acid residues that greatly contribute toward helicase-ligand binding and to allow identification of the ligand pharmacophoric moieties that facilitate these interactions. The generated model thereafter serves as a chemical scaffold that is used to screen through

the ZINC Database via the ZINCPharmer online software [47–49]. The resultant screened compounds were filtered based on toxicity and druggability using ADMET properties and Lipinski's rule of five [50–53].

## ***2.6 Structure-based Virtual Screening***

The top hits generated from screening were subsequently docked to distinguish ligands based on conformational features of the molecules that directly influence its binding affinity to the ATPase pocket of the helicase [54]. Before docking the ligands, Gasteiger partial charges were allocated to the ligands and atom types were defined using the Autodock Graphical user interface [32]. To assist docking compatibility of the ligands, Raccoon software was utilized to convert the files to pdbqt format [54]. The Lamarckian genetic algorithm was used to create different docking poses, thereupon examined to procure binding affinities ( $\text{kcal.mol}^{-1}$ ) [55]. For every ligand, approximately 50 poses were considered and the five ligands, which bound the with highest binding affinities to the NS3 helicase ATPase pocket were selected. The workflow adapted in this study is represented in Figure 2.

## ***2.7 Validation of Docking Approach***

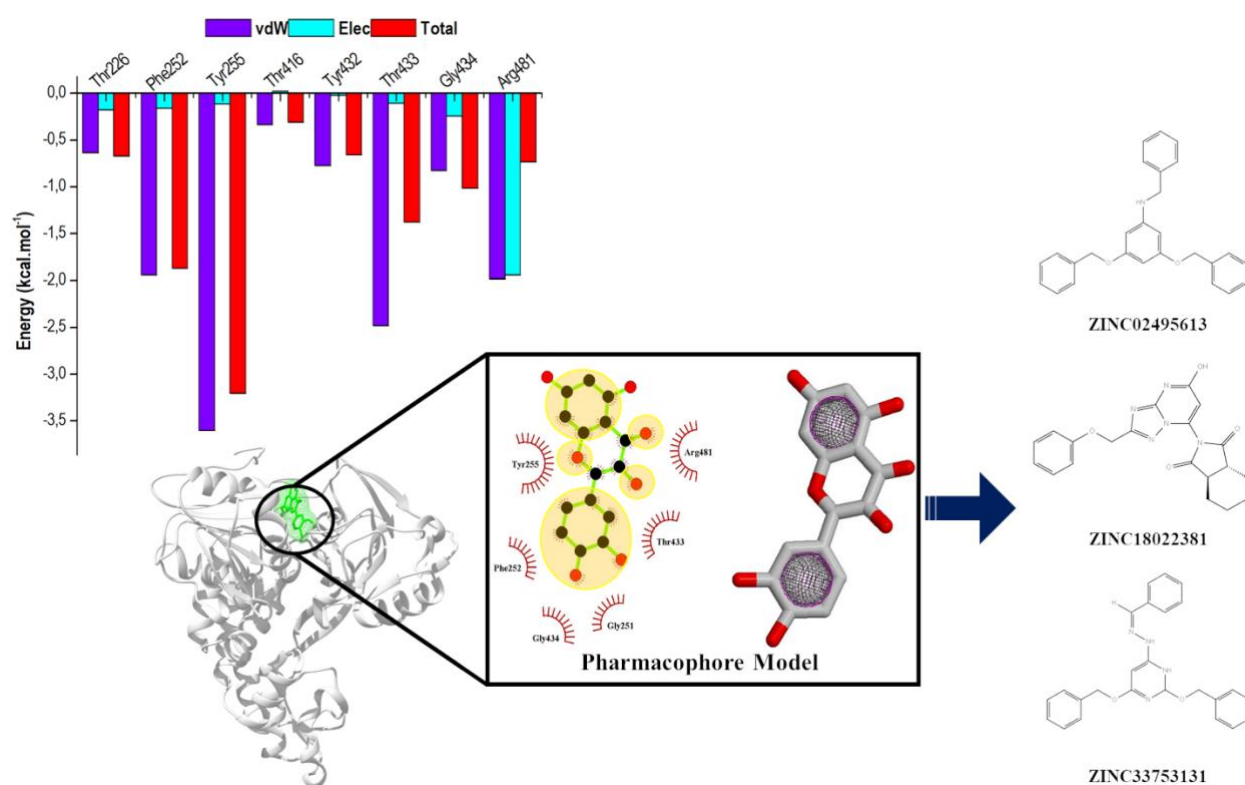
To validate the docking approach employed in this study, the resultant docked quercetin-helicase complex was superimposed onto HCV NS3 helicase protein in complex with a known molecule that binds to the ATP-binding site (Figure S1). Molecular docking generates optimal geometric conformations of a bound ligand in an active pocket, though brief MD simulations do not ensure stability of the complex and bring about disorientation of molecules. Hence, to validate this approach, the three most favorable ligands with highest binding affinities bound to the helicase were subjected to 200ns of MD simulations

## ***3. Results***

### ***3.1 Pharmacophore Model and Virtual Screening***

Over recent years, medicinal chemistry has effectively utilized pharmacophore approaches and is considered one of the most successful and useful strategies in rational drug design. In past attempts during drug research, the binding affinity data was the main source for establishing energy-based pharmacophore; however, flaws such as the emergence of false positives and negatives rendered this approach unreliable. To overcome this problem, the PRED method was employed. The pharmacophoric model presented in Figure 3 was used as a 3D structural template to search the ZINCPharmer database, which yielded a total of three hits. The generated library was then docked using Autodock Vina with the crystal structure of HCV NS3 helicase (PDB: 3O8D). After screening the library, three

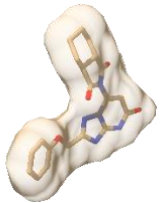
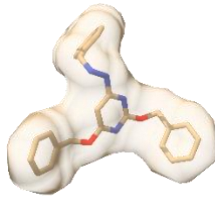
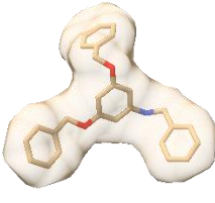
compounds with the highest binding affinities were selected from the library (Table 1). Lipinski's rule of five is based on a set of property values such as the number of hydrogen-bond donors and acceptors, the molecular weight and the *logP*, derived from the FDA approved drugs, which are known to have clinically acceptable ADME properties. Thus, molecules that pass the Lipinski rule are expected to be active in humans after oral administration. It must be noted that all the three hit compounds obeyed Lipinski's rules. The three hit compounds were then subjected to a 200ns MD simulation a studied to precisely determine the ligand–receptor models in the state close to realistic conditions and to further explore the binding modes of the two inhibitors.



**Figure 3:** Pharmacophore model generation from quercetin-helicase complex. The highlighted yellow regions represent the pharmacophoric moieties that were selected for the model, based on residues that contributed the most to the binding affinity. Using ZINCPharmer, three hits were generated from the pharmacophore-based virtual screening process.



**Table 1:** The ID codes, docked 3D-structures and Lipinski's rule of five for the three hits obtained through virtual screening using ZINCPharmer [49,56,57].

Hit	ID	3D-Structure	BA (kcal. mol <sup>-1</sup> )	xLogP	MW (g/mol)	HB <sub>D</sub>	HB <sub>A</sub>
1	ZINC18022381		-8.7	2.60	392.395	0	9
2	ZINC33753131		-8.5	5.39	411.485	2	6
3	ZINC02495613		-8.4	6.42	395.502	1	3

BA: Binding affinity; MW: molecular weight; HB<sub>D</sub>: hydrogen bond donor; HB<sub>A</sub>: hydrogen bond acceptor

### 3.2 Molecular Dynamic Simulations

Molecular dynamics simulations of 200ns and MM/GBSA BFE calculations for the three top-ranked compounds and Quercetin in complex with NS3 helicase, were conducted to validate the accuracy of the binding affinities obtained from the docking calculations.

### 3.3 System Stability

The stability of Quercetin and the top-ranked systems were assessed by calculating the root-mean-square deviation (RMSD) for the respective systems. A plot of the RMSD for the simulated systems can be found in the supplementary section (Figure S2). All systems reached convergence, which validated their stability in the binding site of HCV NS3 helicase.

### 3.4 Molecular Dynamic Simulations and Free Binding Energy Calculations

The MM/GBSA method is widely used to estimate the BFE of an inhibitor at an atomic perspective [58]. The MM/GBSA method was used to evaluate the total binding energy contributions of quercetin and the top hits bound to the NS3 helicase as presented in Table 2.

**Table 2:** Summary of MM/GBSA-based binding free energy contributions of Quercetin and the top-ranked compounds with HCV NS3 helicase.

	Energy components (kcal.mol <sup>-1</sup> )				
	$\Delta E_{vdW}$	$\Delta E_{elec}$	$\Delta G_{gas}$	$\Delta G_{solv}$	$\Delta G_{bind}$
<b>Quercetin</b>	-33.06 ± 4.03	-8.71 ± 5.78	-41.77 ± 7.40	21.92 ± 4.47	-19.85 ± 4.32
<b>ZINC18022381</b>	-38.71 ± 3.75	-16.33 ± 5.03	-55.04 ± 6.41	32.88 ± 4.77	-22.15 ± 3.23
<b>ZINC33753131</b>	-38.51 ± 4.14	-13.30 ± 4.01	-51.82 ± 6.25	26.16 ± 3.15	-25.66 ± 4.59
<b>ZINC02495613</b>	-36.38 ± 3.31	-17.35 ± 6.13	-53.73 ± 6.72	23.13 ± 3.93	-30.61 ± 4.85

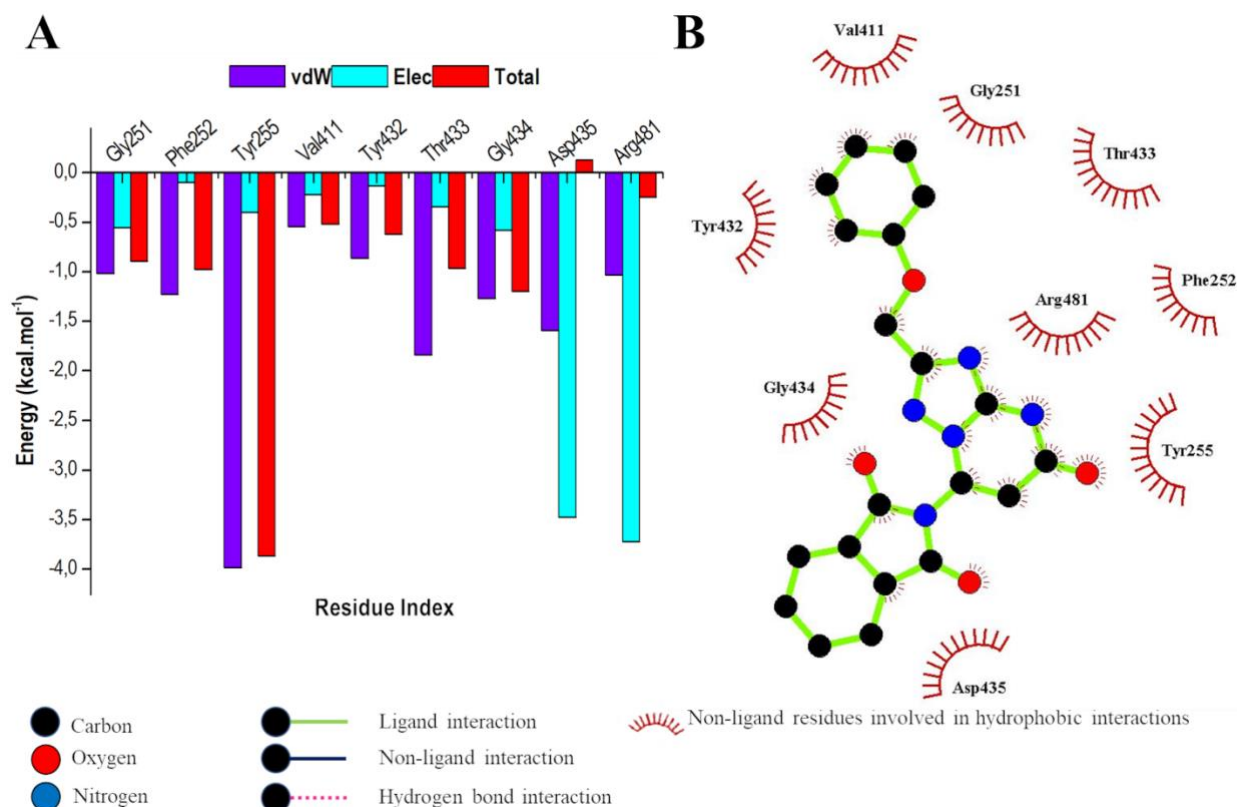
$\Delta E_{VDW}$ : van der Waal;  $\Delta E_{elec}$ : electrostatic;  $\Delta G_{gas}$ : non-polar solvation interaction;  $\Delta G_{solv}$ : solvation free energy;  $\Delta G_{bind}$ : calculated total free binding energy

### 3.5 Per-residue Energy Decomposition

The energetics of the top-ranked hits in complex with NS3 helicase was assessed by computing the per-residue energy contribution to the total BFE of the system. The results obtained are present in Figure 4, 5 and 6.

### 3.5.1 Binding mode analysis of Hit 1

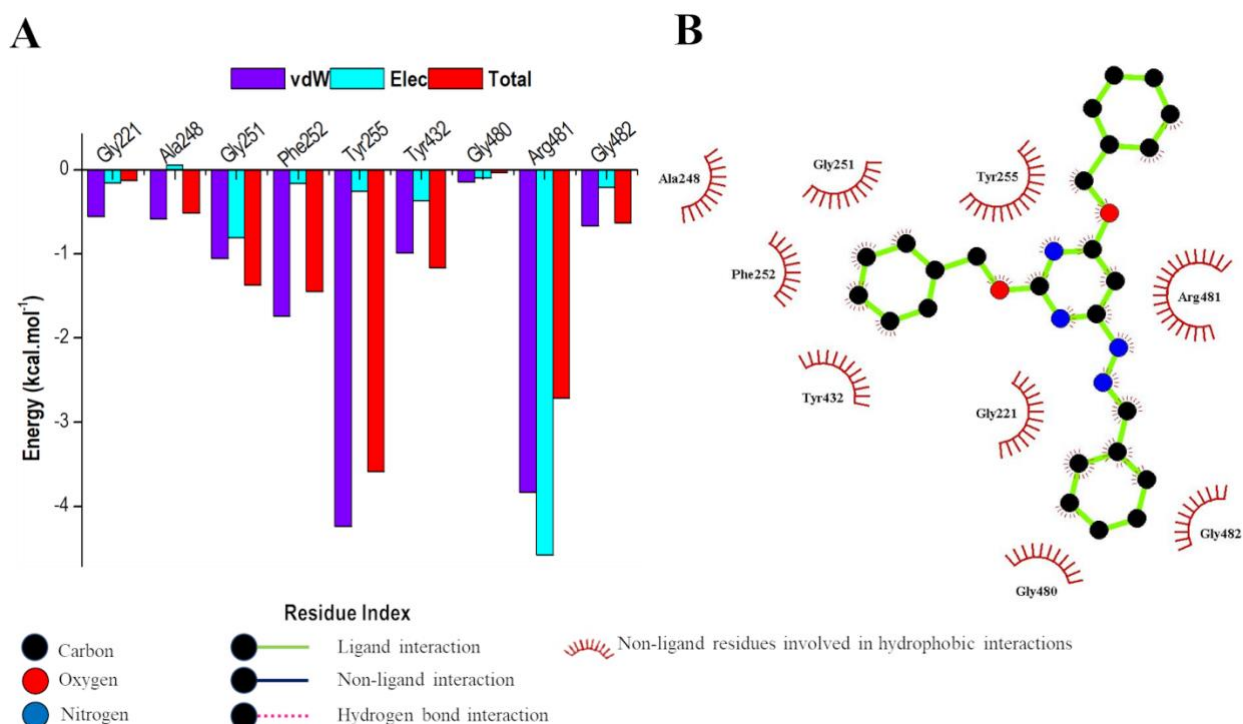
The binding interactions of ZINC18022381 and the per-residue energy contributions of active site residues are presented in Figure 4. ZINC18022381 bound to HCV NS3 helicase binding pocket with a docking score of  $-8.7 \text{ kcal.mol}^{-1}$  and exhibited a higher MM/GBSA BFE ( $-22.15 \text{ kcal.mol}^{-1}$ ) than quercetin ( $-8.1 \text{ kcal.mol}^{-1}$  and  $-19.85 \text{ kcal.mol}^{-1}$ , respectively).



**Figure 4:** (A) The per-residue energy decomposition analysis and (B) ligand interaction plot of ZINC18022381 bound to the ATP-binding site of HCV NS3 helicase.

### 3.5.3 Binding mode analysis of Hit 2

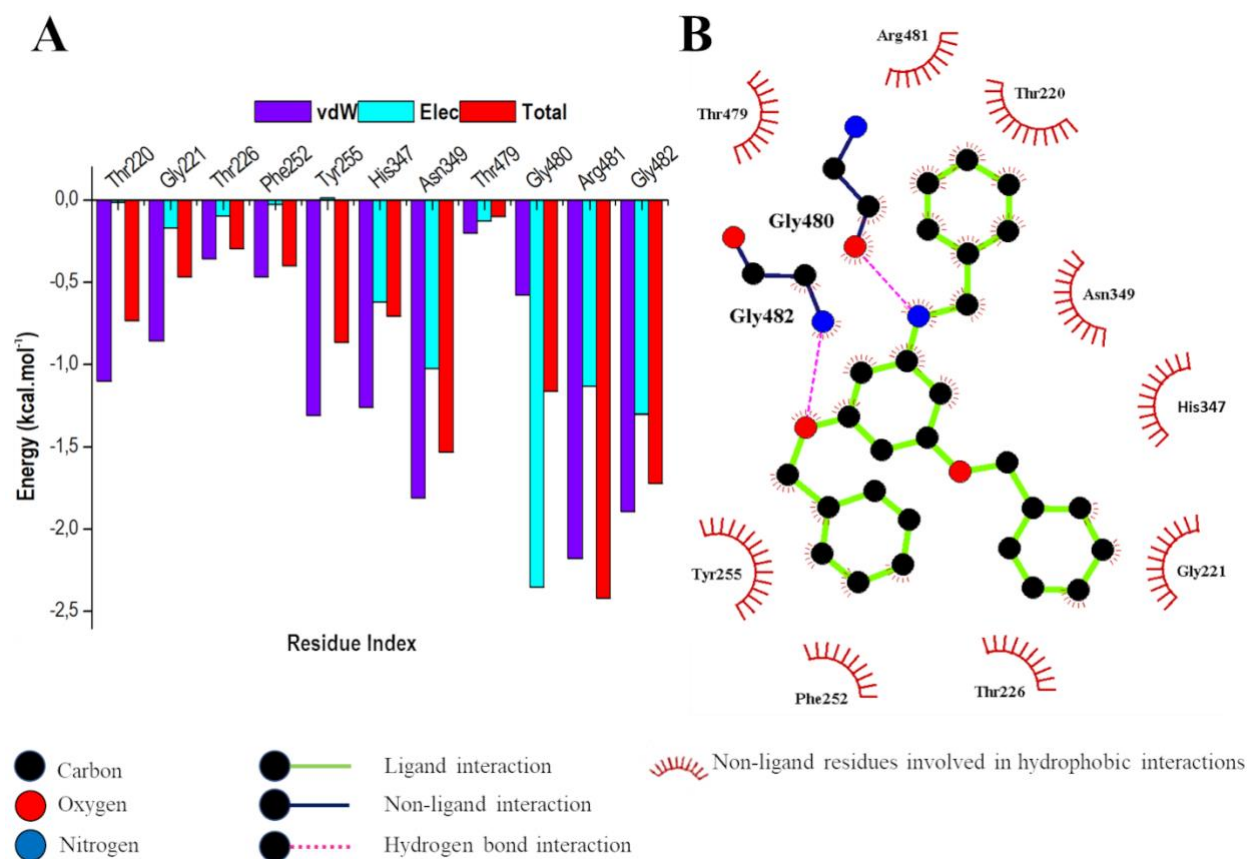
The binding interactions of ZINC33753131 and the per-residue energy contributions of active site residues are presented in Figure 5. ZINC33753131 bound to HCV NS3 helicase binding pocket with a docking score of  $-8.5 \text{ kcal.mol}^{-1}$  and exhibited a higher MM/GBSA BFE ( $-25.66 \text{ kcal.mol}^{-1}$ ) compared to quercetin ( $-8.1 \text{ kcal.mol}^{-1}$  and  $-19.85 \text{ kcal.mol}^{-1}$ , respectively).



**Figure 5:** (A) The per-residue energy decomposition analysis and (B) ligand interaction plot of ZINC33753131 bound to the ATP-binding site of HCV NS3 helicase.

### 3.5.4 Binding mode analysis of Hit 3

The binding interactions of ZINC02495613 and the per-residue energy contributions of active site residues are presented in Figure 6. ZINC02495613 bound to HCV NS3 helicase binding pocket with a docking score of  $-8.4 \text{ kcal.mol}^{-1}$  and exhibited the highest MM/GBSA BFE ( $-30.61 \text{ kcal.mol}^{-1}$ ) compared to quercetin ( $-8.1 \text{ kcal.mol}^{-1}$  and  $-19.85 \text{ kcal.mol}^{-1}$ , respectively) and the other ZINC18022381 and ZINC33753131.



**Figure 6:** (A) The per-residue energy decomposition analysis and (B) ligand interaction plot ZINC02495613 bound to the ATP-binding site of HCV NS3 helicase.

#### 4. Discussion

The highest relative BFE observed in all the systems result from interactions established between the ligand within the active site of the helicase. ZINC02495613 (H3) in complex with NS3 helicase exhibited the highest BFE (-30.61 kcal.mol<sup>-1</sup>) in comparison to quercetin (-19.85 kcal.mol<sup>-1</sup>), ZINC18022381 (H1, -22.15 kcal.mol<sup>-1</sup>) and ZINC33753131 (C3, -25.66 kcal.mol<sup>-1</sup>), as shown in Table 2. From the relevant table, the BFE values suggest that ZINC02495613, ZINC33753131 and ZINC18022381 are more effective potential inhibitor of HCV NS3 compared to quercetin. Intermolecular vdW forces (-36.38 kcal.mol<sup>-1</sup>), electrostatic interactions (-17.35 kcal.mol<sup>-1</sup>) and non-polar solvation interactions (-53.73 kcal.mol<sup>-1</sup>) contributed largely toward the total binding energy of ZINC02495613 to NS3. A polar solvation energy contribution of 23.13 kcal.mol<sup>-1</sup> to the total binding free energy was observed for the ZINC02495613 system. The hydrophobic residues that line the ATP-binding site pocket contribute greatly to the free binding energy of each bound system.

The stability of ZINC18022381 in complex with NS3 exhibited an average RMSD of 2.38 Å extrapolated from a 200ns MD trajectory (Figure S2). The intermolecular interactions formed between

the compound and NS3 binding site residues were examined by calculating the per-residue energy contribution to the total BFE. Figure 4 depicts the per-residue contribution and the ligand interaction diagram of ZINC18022381 bound to the HCV NS3 helicase. Analysis of per-residue energy contribution revealed that significant contributions toward BFE emanated from Tyr255, Thr433, Asp435 and Arg481. The binding site residues that contributed the least energy toward the complex were: Gly251, Phe252, Val411, Tyr432 and Gly434. The values of per-residue energy contribution are found in supplementary, Table S1. The optimum binding conformation adopted by ZINC18022381 was most likely due to vdW interactions from Asp435 and Arg481 and electrostatic interaction from Tyr255.

The stability of ZINC33753131 in complex with NS3 exhibited an average RMSD of 2.27 Å extrapolated from a 200ns MD trajectory (Figure S2). Figure 5 depicts the per-residue contribution and the ligand interaction diagram of ZINC33753131 bound to the HCV NS3 helicase. Analysis of per-residue energy contribution discovered that the following residues contributed the most to the over BFE of the complex: Phe252, Tyr255, Arg481. The binding site residues that contributed minimum energy to the total BRE of the complex included: Gly221, Ala248, Gly251, Tyr432, Gly480 and Gly482. The values of per-residue energy contribution are found in supplementary, Table S2. Therefore, the optimum binding conformation of hit is likely the result of electrostatic energy contribution from Arg481 and vdW interactive forces from Tyr255 and Arg481.

The stability of ZINC02495613 in complex with NS3 exhibited an average RMSD of 2.07 Å estimated from a 200ns MD trajectory (Figure S2). Figure 6 depicts the per-residue contribution and the ligand interaction diagram of ZINC02495613 bound to the HCV NS3 helicase. Analysis of per-residue energy contribution discovered that the following residues contributed the most to the overall BFE of the complex: Asn349, Gly480, Arg481 and Gly482. The binding site residues that contributed the least energy to the total BFE of the complex included: Thr220, Gly221, Thr226, Phe252, Tyr255, His347 and Thr479. The binding mode of ZINC02495613 revealed that ZINC02495613 forms H-bonds with Gly480 and Gly482. The calculated H-bond distance between the nitrogen atom of Gly480 and the oxygen atom of the compound is 2.79 Å while the distance between Gly482 nitrogen atom and the compounds oxygen atom is 3.26 Å. Furthermore, the hydrophobic interactions from the surrounding residues stabilize the conformation of the ligand in the active site. The values of per-residue energy contribution are found in supplementary, Table S3. The formation of hydrogen bonds and generated hydrophobic interactions signifies an optimum binding conformation adopted by hit 3 within the active site.

## ***5. Conclusion***

The HCV remains a global threat and requires continuous research toward its prevention and treatment. Computational approaches allow an in-depth perspective into the underlying structural and dynamical mechanism involved in ligand-enzyme binding, which can be observed at an atomistic level. The utilization of computational techniques accelerates the overall process of drug discovery whilst being cost-efficient. This study acts as an incentive to initiate broader screening of important molecules as HCV helicase inhibitors. The lead compounds identified in this study serve as a promising start for larger scale screenings. The techniques employed in this study were selected for their efficiency, reliability, and for their capacity to aid effective design for lead optimization and potential drug development. The HCV helicase is undoubtedly a very under-developed drug target in literature, evident by the considerably large gap in the manner of which computational studies are performed on potential inhibitors of the helicase. However, recent advancements in the pharmaceutical industry may pave the way for new attempts to discover improved helicase inhibitors. With continued research and on-going advances in this field of study, hopefully the efforts will culminate in the discovery of more compounds for the use in clinical trials and laboratory practices. Further research toward understanding the structural and dynamic mechanisms of HCV enzymes in their natural biological system would be imperative in our pursuit to design and discover enhanced anti-viral inhibitors against this virus.

A total of three compounds, ZINC18022381, ZINC33753131 and ZINC02495613 exhibited similar binding conformation and higher binding affinity compared to quercetin. The stability and docking results of the compounds in complex with NS3 were validated by performing a 200ns MD simulation. All three systems also exhibited a higher binding affinity and binding free energy compared to the weaker inhibitor, quercetin; however, ZINC02495613 stands out as a more effective potential inhibitor of HCV NS3. The hit compounds identified in this study through application of a pharmacophore model, still require biological testing in the future, for the rationalisation of their activity against HCV NS3. The integrated protocol presented in this study can serve as a valuable tool to significantly boost efforts in drug discovery and development processes. It can substantially contribute to the design of promising novel drugs against HCV NS3 helicase and a wide variety of other biological targets, where further experimental investigation of disinterred lead compounds would be beneficial.

### **Future Perspective**

The NS3 helicase plays a pivotal role in HCV replication, however there is an evident lack of FDA approved inhibitors that act against the enzyme, thus the design and development of potential HCV helicase inhibitors should be highly prioritized. Virtual screening is a powerful tool used in pharmaceutical chemistry, it accelerates the discovery process whilst remaining cost-efficient. This is the first study, concerning HCV research, that establishes an alternative approach that may assist

medicinal chemists with the design and discovery of novel inhibitors against HCV helicase and numerous other biological targets. In the future, this study will enhance screening, design and synthesis of novel HCV helicase as well as other viral protein inhibitors, which can then be used in clinical trials evaluating treatment outcomes against HCV.

### Summary Points

- The Hepatitis C virus (HCV) is rapidly becoming a global healthcare and economic burden, affecting more than 3% of the world's inhabitants.
- To date there is no available vaccine for use and the current therapeutic regimes on the market are ineffective due to the onset of system toxicity and drug resistance.
- The HCV RNA helicase plays an important role in viral replication as it is responsible for transcription, translation and splicing of RNA. However, the immense popularity of NS3 protease has greatly overshadowed the potential of RNA helicase as a biological target in drug design research.
- Quercetin is a well-known viral inhibitor of several viruses including those in the *Flavivirus* family but exhibits instability and poor solubility within a physiological medium and is therefore considered a weak inhibitor.
- Hence, we applied an in-house pharmacophore-based virtual screening protocol for identification of potential HCV helicase inhibitors displaying improved pharmacodynamic profiles that is based on highly contributing residues that bind with highest affinity to the weak inhibitor model.
- Through ZINCPharmer, three lead compounds were identified and subjected to 200ns MD simulation. Results obtained from MM/GBSA suggested that all three compounds exhibited similar binding conformation and higher binding affinity in comparison to quercetin, with ZINC02495613 as a noticeably more effective potential inhibitor of HCV NS3 helicase.
- This is a preliminary study that serves as a starting platform to actuate future larger-scaled screening of potential HCV helicase inhibitors and can be integrated in the design, discovery and development of antiviral drugs.

### Acknowledgments

This work was financially supported by the German Academic Exchange Service (DAAD UID:109717) and the National Research Foundation (NRF). Computational support was provided by the Centre for High Performance Computing (CHPC, <http://www.chpc.ac.za>).



**Disclosure**

No potential conflict of interest was reported by the authors. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

## References

1. Karoney MJ, Siika AM, Karoney MJ. Hepatitis C virus (HCV) infection in Africa: a review. *Pan African Med. Journal. Pan African Med. J.* [Internet]. 141444, 44–1937 (2013). Available from: <http://www.panafrican-med-journal.com/content/article/14/44/full/>.
2. Simmonds P, Becher P, Bukh J, *et al.* ICTV Virus Taxonomy Profile : Flaviviridae. *J. Gen. Virol.* (2017).
3. Petruzzello A, Marigliano S, Loquercio G, Cozzolino A, Cacciapuoti C. Global epidemiology of hepatitis C virus infection: An up-date of the distribution and circulation of hepatitis C virus genotypes. *World J. Gastroenterol.* 22(34), 7824–7840 (2016).
4. Kazakov T, Yang F, Ramanathan HN, Kohlway A, Diamond MS, Lindenbach BD. Hepatitis C Virus RNA Replication Depends on Specific Cis- and Trans-Acting Activities of Viral Nonstructural Proteins. *PLoS Pathog.* 11(4), e1004817 (2015).
5. Saw WG, Pan A, Subramanian Manimekalai MS, Grüber G. Structural features of Zika virus non-structural proteins 3 and -5 and its individual domains in solution as well as insights into NS3 inhibition. *Antiviral Res.* 141(1), 73–90 (2017).
6. Jain R, Coloma J, García-Sastre A, Aggarwal AK. Structure of the NS3 helicase from Zika virus. *Nat. Struct. Mol. Biol.* 23(8), 752–754 (2016).
7. Pathak N, Lai ML, Chen WY, Hsieh BW, Yu GY, Yang JM. Pharmacophore anchor models of flaviviral NS3 proteases lead to drug repurposing for DENV infection. *BMC Bioinformatics.* (2017).
8. Raney KD, Sharma SD, Moustafa IM, Cameron CE. Hepatitis C virus non-structural protein 3 (HCV NS3): A multifunctional antiviral target. *J. Biol. Chem.* 285(30), 22725–22731 (2010).
9. Kwong AD, Kauffman RS, Hurter P, Mueller P. Discovery and development of telaprevir: an NS3-4A protease inhibitor for treating genotype 1 chronic hepatitis C virus. *Nat. Biotechnol.* [Internet]. 29(11), 993–1003 (2011). Available from: <http://dx.doi.org/10.1038/nbt.2020%5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/22068541>.
10. Chen KX, Njoroge FG. The journey to the discovery of boceprevir: An NS3-NS4 HCV protease inhibitor for the treatment of chronic hepatitis C. *Prog. Med. Chem.* , 1–36 (2010).
11. Rosenquist Å, Samuelsson B, Johansson PO, *et al.* Discovery and development of simeprevir (TMC435), a HCV NS3/4A protease inhibitor. *J. Med. Chem.* (2014).

12. Keating GM. Elbasvir/Grazoprevir: First Global Approval. *Drugs*. (2016).
13. Markham A, Keam SJ. Danoprevir: First Global Approval. *Drugs*. (2018).
14. Venkatesan A, Rambabu M, Jayanthi S, Febin Prabhu Dass J. Pharmacophore feature prediction and molecular docking approach to identify novel anti-HCV protease inhibitors. *J. Cell. Biochem.* (2018).
15. Shadrick WR, Ndjomou J, Kolli R, Mukherjee S, Hanson AM, Frick DN. Discovering new medicines targeting helicases: challenges and recent progress. *J. Biomol. Screen.* 18(7), 761–781 (2013).
16. Fatima K, Mathew S, Suhail M, *et al.* Docking studies of Pakistani HCV NS3 helicase: A possible antiviral drug target. *PLoS One*. (2014).
17. Li Y, Yao J, Han C, *et al.* Quercetin, inflammation and immunity. *Nutrients*. 8(3), 167 (2016).
18. Pan A, Saw WG, Manimekalai S, *et al.* Structural features of NS3 of Dengue virus serotypes 2 and 4 in solution and insight into RNA binding and the inhibitory role of quercetin. *Acta Crystallogr. Sect. D Struct. Biol.* 73(5), 402–419 (2017).
19. Nathiya S, Durga M, Devasena T. Quercetin, encapsulated quercetin and its application- A review. *Int. J. Pharm. Pharm. Sci.* (2014).
20. Yu W, MacKerell ADJ. Computer-Aided Drug Design Methods. *Methods Mol. Biol.* 1520, 85–106 (2017).
21. Das PS, Saha P, Abdul APJ. A Review on Computer Aided Drug Design In Drug Discovery. *World J. Pharm. Pharm. Sci.* 6(7), 279–291 (2017).
22. Kore PP, Mutha MM, Antre R V., Oswal RJ, Kshirsagar SS. Computer-Aided Drug Design: An Innovative Tool for Modeling. *J. Med. Chem.* 2(1), 139–148 (2012).
23. Lionta E, Spyrou G, Vassilatis DK, Cournia Z. Structure-based virtual screening for drug discovery: principles, applications and recent advances. *Curr. Top. Med. Chem.* 14(16), 1923–1938 (2014).
24. Appiah-Kubi P, Soliman ME. Hybrid receptor-bound/MM-GBSA-per-residue energy-based pharmacophore modelling: enhanced approach for identification of selective LTA4H inhibitors as potential anti-inflammatory drugs. *Cell Biochem. Biophys.* 75(1), 35–48 (2017).
25. Berinyuy E, Soliman MES. Identification of Novel Potential gp120 of HIV-1 Antagonist Using Per-Residue Energy Contribution-Based Pharmacophore modelling. *Interdiscip. Sci. Comput. Life Sci.* 9(3), 406–418 (2017).

26. N. Mhlongo N, E. S. Soliman M. Binding Free Energy-Based Footprint Pharmacophore Model to Enhance Virtual Screening and Drug Discovery: A Case on Glycosidases as Anti-influenza Drug Targets. *Lett. Drug Des. Discov.* 13(10), 1033–1046 (2016).
27. Cele FN, Ramesh M, Soliman MES. Per-residue energy decomposition pharmacophore model to enhance virtual screening in drug discovery: A study for identification of reverse transcriptase inhibitors as potential anti-HIV agents. *Drug Des. Devel. Ther.* 10, 1365–1377 (2016).
28. Islam MA, Pillay TS. Exploration of the structural requirements of HIV-protease inhibitors using pharmacophore, virtual screening and molecular docking approaches for lead identification. *J. Mol. Graph. Model.* 56, 20–30 (2015).
29. Appleby TC, Anderson R, Fedorova O, *et al.* Visualizing ATP-dependent RNA translocation by the NS3 helicase from HCV. *J. Mol. Biol.* (2011).
30. Hanwell MD, Curtis DE, Lonie DC, Vandermeersch T, Zurek E, Hutchison GR. Avogadro: An advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminform.* 4(8), 17 (2012).
31. Yang Z, Lasker K, Schneidman-Duhovny D, *et al.* UCSF Chimera, MODELLER, and IMP: an Integrated Modeling System. *J. Struct. Biol.* 179(3), 269–278 (2012).
32. Trott O, Olson AJ. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J. Comput. Chem.* 31(2), 455–61 (2010).
33. Pettersen EF, Goddard TD, Huang CC, *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25(13), 1605–1612 (2004).
34. Fogolari F, Corazza A, Toppo S, *et al.* Studying interactions by molecular dynamics simulations at high concentration. *J. Biomed. Biotechnol.* [Internet]. 2012, 303190 (2012). Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3303702&tool=pmcentrez&render type=abstract>.
35. Case DA, Babin V, Berryman J, *et al.* Amber 14. (2014).
36. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* 11(8), 3696–3713 (2015).
37. Ramharack P, Oguntade S, Soliman MESS. Delving into Zika virus structural dynamics – a closer look at NS3 helicase loop flexibility and its role in drug discovery. *RSC Adv.* [Internet]. 7(36), 22133–22144 (2017). Available from: <http://xlink.rsc.org/?DOI=C7RA01376K>.

38. Mhlongo NN, Ebrahim M, Skelton AA, Kruger HG, Williams IH, Soliman MES. Dynamics of the thumb-finger regions in a GH11 xylanase *Bacillus circulans*: comparison between the Michaelis and covalent intermediate†. *RSC Adv.* [Internet]. 5(100), 82381–82394 (2015). Available from: <http://dx.doi.org/10.1039/C5RA16836H>.
39. Gonnet P. P-SHAKE: A quadratically convergent SHAKE in  $O(n^2)$ . *J. Comput. Phys.* 220(2), 740–750 (2007).
40. Le Grand S, Götz AW, Walker RC. SPFP: Speed without compromise—A mixed precision model for GPU accelerated molecular dynamics simulations. *Comput. Phys. Commun.* 184(2), 374–380 (2013).
41. Gapsys V, Michielssens S, Peters JH, de Groot BL, Leonov H. Calculation of binding free energies. *Methods Mol. Biol.* 1215, 173–209 (2015).
42. Schauerl M, Czodrowski P, Fuchs JE, *et al.* Binding Pose Flip Explained via Enthalpic and Entropic Contributions. *J. Chem. Inf. Model.* 57(2), 345–354 (2017).
43. Agoni C, Ramharack P, Soliman MES. Co-inhibition as a strategic therapeutic approach to overcome rifampin resistance in tuberculosis therapy: Atomistic insights. *Future Med. Chem.* 10(14), 1665–1675 (2018).
44. Hayes JM, Archontis G. MM-GB ( PB ) SA Calculations of Protein-Ligand Binding Free Energies. In: *Molecular Dynamics – Studies of Synthetic and Biological Macromolecules*. InTech (2012).
45. Hou T, Wang J, Li Y, Wang W. Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. *J. Chem. Inf. Model.* 51(1), 69–82 (2011).
46. Zhang X, Perez-Sanchez H, C. Lightstone F. A Comprehensive Docking and MM/GBSA Rescoring Study of Ligand Recognition upon Binding Antithrombin. *Curr. Top. Med. Chem.* [Internet]. 17(14), 1631–1639 (2017). Available from: <http://www.eurekaselect.com/openurl/content.php?genre=article&issn=1568-0266&volume=17&issue=14&spage=1631>.
47. Irwin JJ, Shoichet BK. ZINC – a free database of commercially available compounds for virtual screeningNo Title. *Journal Chem. Inf. Model.* 45(1), 177–182 (2005).
48. Koes DR, Camacho CJ. ZINCPharmer: Pharmacophore search of the ZINC database. *Nucleic Acids Res.* 40(W1), 409–414 (2012).
49. Awale M, Reymond J. A multi-fingerprint browser for the ZINC database. *Nucleic Acids Res.*

- 42(W1), W234–W239 (2014).
50. Lipinski CA. Lead- and drug-like compounds: The rule-of-five revolution. *Drug Discov. Today Technol.* 1(4), 337–341 (2004).
  51. Tian S, Wang J, Li Y, Li D, Xu L. The application of in silico drug-likeness predictions in pharmaceutical research. *Adv. Drug Deliv. Rev.* 86, 2–10 (2015).
  52. Arnott JA, Planey SL. The influence of lipophilicity in drug discovery and design. *Expert Opin. Drug Discov.* 7(10), 863–875 (2012).
  53. Terfloth L, Spycher S, Gasteiger J. Drug Discovery: An Overview. In: *Applied Chemoinformatics: Achievements and Future Opportunities*. Engel T, Gasteiger J (Eds.). . Wiley, 165–194 (2018).
  54. Forli S, Huey R, Pique ME, Sanner MF, Goodsell DS, Olson AJ. Computational protein–ligand docking and virtual drug screening with the AutoDock suite. *Nat. Protoc.* 11(1), 905–919 (2016).
  55. Fuhrmann J, Rurainski A, Lenhof H, Neumann D. A new Lamarckian genetic algorithm for flexible ligand-receptor docking. *J. Comput. Chem. Chem.* 31(9), 1911–1918 (2010).
  56. Koes DR, Camacho CJ. ZINCPharmer: Pharmacophore search of the ZINC database. *Nucleic Acids Res.* 40(W1) (2012).
  57. Irwin JJ, Sterling T, Mysinger MM, Bolstad ES, Coleman RG. ZINC: A free tool to discover chemistry for biology. *J. Chem. Inf. Model.* 52(7), 1757–1768 (2012).
  58. Genheden S, Ryde U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin. Drug Discov.* [Internet]. 10(5), 449–461 (2015). Available from: <http://www.tandfonline.com/doi/full/10.1517/17460441.2015.1032936>.

## CHAPTER 7

### 7.1 Conclusion

Hepatitis C Virus (HCV) is a major causative agent of liver disease, worldwide. In the last ten years, basic understanding of HCV molecular virology and life cycle has advanced significantly with the use of HCV cell culture and replication system models coupled with enhancements in computer-aided drug design. This evolution in science has greatly aided rational drug design efforts, culminating in the generation of numerous antiviral drugs. However, current HCV therapeutics are deemed ineffective amongst the majority of patients due to viral genetic diversity, lack of specificity and the onset of adverse effects. This led to the re-emergence of covalent inhibition as a popular topic in the pharma industry. Scientists have taken the strategy one step further and proposed that covalent modifications of viral therapeutic targets through conserved cysteine amino acid residues have the potential to substantially contribute toward drug's attaining viral selectivity. Selective covalent inhibition in HCV is a highly neglected topic in science and little is known regarding its effect and influence on the virus.

This study sought to devise a criterion for the efficient implementation of selective covalent inhibition in HCV drug design by introducing a technical guideline or “road map” for the design and discovery of optimized potential selective covalent inhibitors against HCV NS3/4A protease and other viral targets. This study also investigated selective covalent inhibition in HCV through one of the key viral enzymes, NS5B RNA-dependent RNA polymerase. This mode of inhibition induced reorganization of key protein elements required for RNA transcription led by modifications in enzyme flexibility, correlated dynamics and intermolecular bonding, ultimately compromising the enzyme's overall structural integrity. These insights into the mechanism of selective covalent inhibition offers significant understanding that will promote identification and discovery of potent, selective inhibitors against HCV with the potential to overcome challenges that have given way to drug resistance.

This study further, shed light on HCV RNA helicase as a prospective target in drug design and proposed a preliminary investigation into the discovery of potential inhibitors of the helicase. The HCV helicase was subjected to “per-residue energy decomposition pharmacophore” virtual screening to identify favourable molecules that are more effective than the experimentally tested inhibitor. The results reported in this study will undoubtedly enhance future endeavours in designing and developing potentially effective inhibitors against HCV.

Overall, this study offers a valuable insight into the mechanism of selective covalent inhibition, uncovering previously neglected viral target and identifying possible antiviral drugs through molecular modelling and computer-aided drug design.

## **7.2 Future Perspective**

The computational approaches and parameters adopted in the present study are cost-efficient and powerful tools that can be effectively implemented in the process of drug design and discovery. The work presented in this study reveals a suitable and reliable technical guideline that may be used to initiate the design, discovery and development of optimized potential selective covalent HCV inhibitors.

The work in this study can also facilitate further understanding of the mechanism underlying selective covalent inhibitors in use or development and their impact in a variety of diseases. Very little is understood and known about selective covalent inhibition due to focus being directed at other areas, for example, non-covalent inhibitors. To gain the true benefits of this mode of inhibition, the pharma industry must undergo a paradigm shift toward selective inhibition to advance and enhance current HCV therapies.

To further examine the simulation validity and proposed parameters of the systems studied, the MD simulation can be carried out in triplicate with a longer simulation period. Performing system simulations in triplicate will aid in result validation as well as providing further insight and understanding into the mechanism of selective covalent inhibition.

The potential inhibitors of the study have exhibited favourable binding energies and promising protein-drug interactions and therefore maybe utilized as the lead compounds in anti-HCV therapy. However, future biological testing of these compounds is still required to verify and validate this *in silico* approach.



## APPENDICES

### Appendix A

Biomedical Research Ethics Approval

### Appendix B

Letitia Shunmugam, Pritika Ramharack and Mahmoud E.S. Soliman (2017). Road Map for the Structure-Based Design of Selective Covalent HCV NS3/4A Protease Inhibitors. *The Protein Journal*, 36(5), 397-406. (*Published*)

### Appendix C

Letitia Shunmugam and Mahmoud E.S. Soliman (2018). Targeting HCV Polymerase: A Structural and Dynamic Perspective into the Mechanism of Selective Covalent Inhibition. *RSC Advances*, 8(73), 42210-42222 (*Published*)

### Appendix D

Letitia Shunmugam, Nikita Devnarain and Mahmoud E.S. Soliman (2018). Enhanced Pharmacophore-based Virtual Screening Approach in the Discovery of Potential Inhibitors of HCV NS3 Helicase, *Future Virology*, Manuscript ID: FVL-2019-0006 (*Submitted*)