MOLECULAR DOCKING AND MOLECULAR DYNAMICS STUDY ON POTENTIAL EBOLA MATRIX PROTEIN VP40 INHIBITORS

BY

MOHAMAD ARIFF BIN MOHAMAD YUSSOFF

A thesis submitted in fulfilment of the requirement for the degree of Master of Science (Biotechnology)

Kulliyyah of Science
International Islamic University Malaysia

NOVEMBER 2019

ABSTRACT

It is well known that Ebola virus (EBOV) causes severe haemorrhagic fever with high fatality rate. The biggest Ebola outbreak in 2014 has caused at least 11,323 deaths with more than 28,000 cases have been reported. Studies have demonstrated that EBOV matrix protein known as VP40 is crucial in the early infection stage to facilitate the transcription of the viral gene through association with ssRNA in specific manner. Two residues namely Phe125 and Arg134 were found to be important in mediating the VP40-RNA interaction. Thus, blocking this VP40-RNA interaction could interrupt EBOV life cycle in the host cell. This study aims to identify and optimise ligands that can potentially block the RNA binding site of VP40. A total of 42 previously studied ligands from literature were simulated against the RNA binding site using AutoDock Vina. The top ten ligands were used as templates for similarity search in ZINC databases using USRCAT followed by structure-based virtual screening at the RNA binding site and molecular dynamics simulation. The ADME properties of the compounds were predicted computationally and the binding free energy of the complex was calculated using molecular-mechanics Poisson Boltzmann surface area (MM-PBSA) method. Our results showed that Q-88 (ZINC ID: 1342431) turned out to be the best-docked compound with binding free energy of -97.27 kJ/mol. However, this compound gave unsatisfactory ADME properties by violating two Ghose's rule and has low GI absorption. Substituting the sulphur (S) atom with oxygen (O) in the backbone structure of Q-88 eliminated the drug-likeness violation and improved GI absorption prediction with similar binding free energy (-97.097 kJ/mol). This finding shed light on binding energies and potential modification of previously tested compounds against Ebola VP40, which could be useful to design more potent and drug-like VP40 inhibitors.

خلاصة البحث

إن من المعروف أن فيروس الإيبولا (EBOV) يسبب الحمى النزفية الحادة وبمعدل وفيات مرتفعة. قد تسبب تفشي فيروس الإيبولا في عام 2014 في وفاة 11323 شخصًا على الأقل، مع بلاغات عن أكثر من 28000 حالة. أظهرت الدراسات أن بروتين مصفوفة فيروس الإيبولا المعروف باسم VP40 بروتين مهم في المرحلة المبكرة للعدوى لتسهيل نسخ الجين الفيروسي من خلال الارتباط مع الحمض النووي الريبوزي ذي الظفيرة الواحدة (ssRNA) بطريقة محددة. يعتبر Phe125 و Arg134 من البقايا المهمة في التوسط في تفاعل مع VP40 و RNA. وبالتالي قد يؤدي وقف تفاعل VP40 و RNA إلى وقف دورة حياة EBOV في الخلية المضيفة. هدفت هذه الدراسة إلى تحديد وتحسين الروابط التي من المحتمل أن تحجب موقع ربط الحمض النووي الريبوزي في VP40. تم محاكاة 42 من الروابط سبق دراستها من المؤلفات ضد موقع ربط الحمض النووي الريبوزي باستخدام برنامج AutoDock Vina. تم استخدام الروابط العشرة الأولى كقوالب للبحث عن المتشابهات في قواعد بيانات ZINC باستخدام USRCAT متبوعة بالفحص الظاهري الهيكلي في موقع ربط RNA ومحاكاة الديناميات الجزيئية. تم التنبؤ بخصائص ADME للمركبات حسابياً وحُسِبَت طاقة الارتباط الحرة للمجمع باستخدام طريقة مساحة السطح الميكانيكية الجزيئية لبويزون بولتزمان (MM-PBSA). أظهرت نتائجنا أن Q-88 (ZINC ID: 1342431) كان المركب الأفضل ربطا بطاقة ارتباط حرة بلغت -97.27 كيلو جول/مول. ومع ذلك فقد أعطى هذا المركب خصائص ADME غير مرضية وذلك بانتهاكه لإثنين من قوانين Ghose وبسبب قيمة امتصاص GI المنخفض. أدى استبدال ذرة الكبريت (S) بالأكسجين (O) في بنية العمود الفقري لـ Q-88 إلى إزالة انتهاك مدى التشابه للأدوية وتحسين التنبؤ بامتصاص GI مع تقارب ربط مماثل (-97.097 كيلو جول/مول). تلقي هذه النتيجة الضوء على طاقات الربط والتعديل المحتمل للمركبات التي تم اختبارها سابقًا ضد VP40الخاص بالإيبولا، والتي قد تكون مفيدة لتصميم مثبطات VP40 أكثر فاعلية وبتشابه أكبر للأدوية.

APPROVAL PAGE

I certify that I have supervised and read this study and that in my opinion, it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a thesis for the degree of Master of Science (Biotechnology).

……..
Khairul Bariyyah Abd Halim
Supervisor

……..
Azzmer Azzar Abdul Hamid
Co-Supervisor

……..
Shafida Abdul Hamid
Co-Supervisor

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a thesis for the degree of Master of Science (Biotechnology).

……..
Zalikha Ibrahim
Internal Examiner

……..
Siti Azma Jusoh
External Examiner

This thesis was submitted to the Department of Biotechnology and is accepted as a fulfilment of the requirement for the degree of Master of Science (Biotechnology).

……..
Mardiana Mohd Ashaari
Head, Department of Biotechnology

This thesis was submitted to the Kulliyyah of Science and is accepted as a fulfilment of the requirement for the degree of Master of Science (Biotechnology).

……..
Shahbudin Saad
Dean, Kulliyyah of Science

DECLARATION

I hereby declare that this thesis is the result of my own investigations, except where otherwise stated. I also declare that it has not been previously or concurrently submitted as a whole for any other degrees at IIUM or other institutions.

Mohamad Ariff bin Mohamad Yussoff

Signature ..... Date .....

INTERNATIONAL ISLAMIC UNIVERSITY MALAYSIA

DECLARATION OF COPYRIGHT AND AFFIRMATION OF FAIR USE OF UNPUBLISHED RESEARCH

MOLECULAR DOCKING AND MOLECULAR DYNAMICS STUDY ON POTENTIAL EBOLA MATRIX PROTEIN VP40 INHIBITORS

I declare that the copyright holders of this thesis are jointly owned by the student and IIUM.

By signing this form, I acknowledged that I have read and understand the IIUM Intellectual Property Right and Commercialization policy.

Affirmed by Mohamad Ariff bin Mohamad Yussoff

……..…….. ……..
Signature Date

ACKNOWLEDGEMENTS

In the name of Allah, the Most Gracious and the Most Merciful. Praise be to Allah for the mercy He has bestowed upon me in completing writing this thesis.

I would like to express my highest gratitude to my supervisor, Dr. Khairul Bariyyah Abd Halim for helping, guiding, advising, inspiring and giving opinions and recommendations from the beginning until the end of this project and in completing my thesis writing. I also like to express my gratitude to my co-supervisors, Dr Azzmer Azzar Abdul Hamid and Dr Shafida Abdul Hamid that helped and guided me as well as giving thoughts to complete my project and thesis writing. This supervisory team is indeed one of the best team that guide me along the journey of this study.

Special thanks to my laboratory team, Br. Mohammad Hakim and Br. Muhammad Hasanuddin for giving their fullest teamwork, cooperation, knowledge and ideas in helping me completing this project.

Lastly, I would like to express my deepest gratitude to my beloved parents, Mohamad Yussoff Ali and Masayu @ Hariati Mukhtar for their encouragement and support in completing this Master study. Thank you for everything.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS/UNITS

Å	angstrom
ns	nanosecond
t	time
α	alpha
β	beta
kcal	kilocalorie
kJ	kilojoule
mol	mole

LIST OF ABBREVIATIONS

EBOV	Ebola virus
EVD	Ebola virus disease
RNA	Ribonucleic acid
WHO	World Health Organization
VP30	Viral protein 30
VP35	Viral protein 35
VP24	Viral protein 24
VP40	Viral protein 40
NC	Nucleocapsid
NP	Nucleoprotein
GP	Glycoprotein
L	L polymerase
VLPs	Virus like particles
PIP_2	Phosphatidylinositol 4,5-bisphosphate
PS	Phosphatidylserine
FDA	Food and Drug Administration
ADME	Absorption, distribution, metabolism, excretion
ADMET	Absorption, distribution, metabolism, excretion, toxicity
MD	Molecular dynamics
SUDV	Sudan virus
BDBV	Bundibugyo virus
RESTV	Reston virus
TAFV	Taï Forest virus
DRC	Democratic Republic of Congo
USA	United States of America
ssRNA	Single stranded ribonucleic acid
RNP	Ribonucleoprotein
ORF	Open reading frame
ESCRT	Endosomal Sorting Complex Required for Transport
NEDD4	Neural precursor cell expressed developmentally down-regulated protein 4
TSG101	Tumour susceptibility gene 101
MVB	Multivesicular body trafficking
SRSs	Suppressor of RNA silencing
miRNA	microRNA
RNAi	RNA interference
TRBP	Transactivation response element RNA-binding protein
PACT	Protein activator of the interferon-induced protein kinase
HIV	Human immunodeficiency virus
HIV-GAG	HIV-group specific antigen
TCM	Traditional Chinese Medicine
CADD	Computer-aided drug design

PDB	Protein Data Bank
3D	Three dimensional
QSAR	Qualitative structure-activity relationship
2D	Two dimensional
USR-VS	Ultrafast Shape Recognition – Virtual Screening
USRCAT	Ultrafast Shape Recognition with CREDO Atom Types
WDI	World Drug Index
ATB	Automated Topology Builder
GROMACS	Groningen Machine for Chemical Simulation
SPC/E	Extended simple point charge
MM-PBSA	Molecular mechanics Poisson-Boltzmann surface area
PSA	Polar surface area
GI	Gastrointestinal
BBB	Blood brain barrier
P-gp	Permeability glycoprotein
CNS	Central nervous system
PAINS	Pan assay interference compounds
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation
Ala	Alanine
Arg	Arginine
Asn	Asparagine
Gln	Glutamine
Gly	Glycine
His	Histidine
Ile	Isoleucine
Leu	Leucine
Lys	Lysine
Phe	Phenylalanine
Pro	Proline
Thr	Threonine
Tyr	Tyrosine

CHAPTER ONE

INTRODUCTION

1.1 BACKGROUND OF THE STUDY

Ebola virus disease (EVD) is a disease that caused by the Ebola virus (EBOV). It is formerly known as Ebola haemorrhagic fever, a severe disease that was first appeared in 1976 in two concurrent outbreaks; one was in Nzara, South Sudan and the other one was in Yambuku, Democratic Republic of Congo (WHO, 1976). EVD is transmitted to people from wild animals and then spreads in the human population through human-to-human transmission. In 2014, the disease appeared in West Africa and is known as the largest and most complex Ebola outbreak compared to the previous outbreak. During that time, EVD has been spread between countries including Liberia, Nigeria and USA (WHO, 2017).

According to Feldmann & Geisbert (2011), EBOV is a lipid-enveloped filamentous virus that affects human and non-human primates causing severe haemorrhagic fever with the fatality rate up to 90%. EBOV harbours a negative-sense RNA genome that encodes seven proteins namely nucleoprotein (NP), viral protein 30 (VP30), viral protein 35 (VP35), L-polymerase (L) protein, viral protein 24 (VP24), viral protein 40 (VP40) and transmembrane glycoprotein (GP) (Figure 1.1). NP, VP30, VP35 and L protein constitute the nucleocapsid (NC) while GP is accounted for the entry of virions into the host cell. Meanwhile, VP24 is a minor protein which is also crucial for the assembly of NC and VP40 is the viral matrix protein that regulates viral budding and NC recruitment, as well as virus structure and stability (Booth, Rabb, & Beniac, 2013; Olejnik, Ryabchikova, Corley, & Mühlberger, 2011).

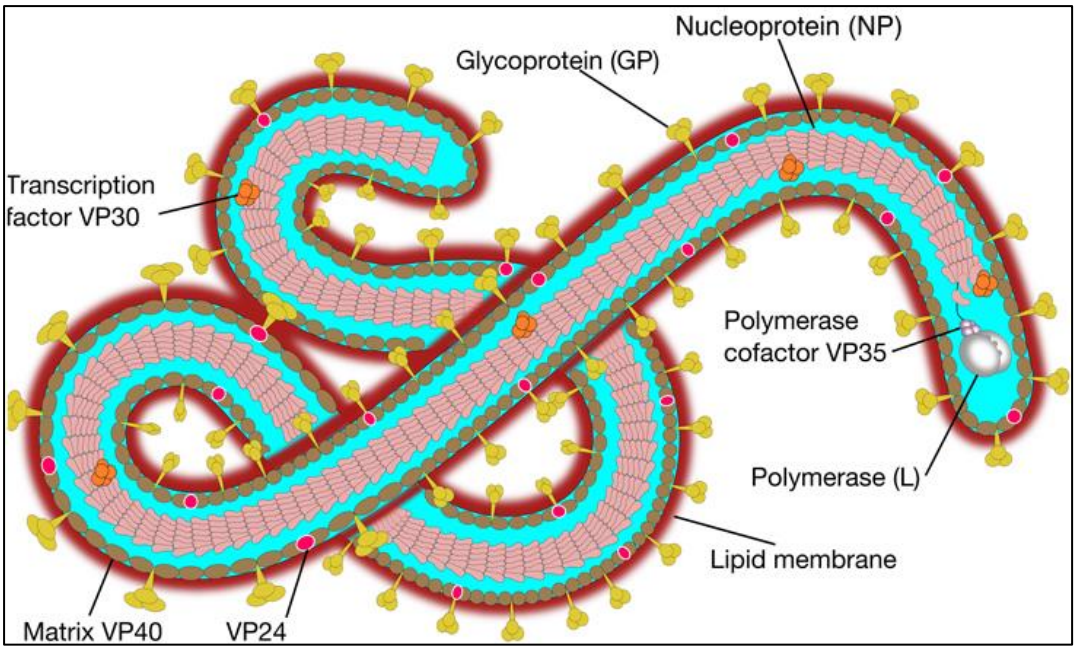

Figure 1.1 Structure of Ebola Virus with Its Encoded Protein. Adapted from SIB SWISS Institute of Bioinformatics (2014).

Jasenosky et al. (2001) stated the importance of VP40 in EBOV which appears equivalent to matrix proteins of other viruses that are able to induce virus-like particles (VLPs) formation. They also found that VP40 is able to bind to cellular membranes in a hydrophobic manner. Another study by Dessen et al. (2000) showed that the X-ray crystal structure of VP40 from EBOV at 2.0 Å resolution consisted of two domains with unique folds. The C-terminal domain has large hydrophobic patches that may involve in the interaction with lipid bilayers for membrane binding while N-terminal domain consists of six antiparallel β-strands and important for VP40 oligomerisation.

Many studies have been carried out to understand more about the interaction of VP40 with the plasma membrane. VP40 has been reported to be associated with negatively charged lipids such as phosphatidylinositol 4,5-bisphosphate (PIP_2) and phosphatidylserine (PS) (Adu-Gyamfi et al., 2015; K. A. Johnson et al., 2016). It has

been hypothesised that the VP40 dimers undergo major structural rearrangement once it interacts with the negatively charged lipids at the plasma membrane and oligomerise into hexameric structures (Bornholdt et al., 2013; Soni et al., 2013). Recent computational study showed that VP40 interact with the negatively charged membrane via two flexible loops that consist of several lysine residues (GC, Gerstman, & Chapagain, 2017).

Although VP40 is important for membrane interaction, this protein is also shown to be responsible for viral replication by involving in RNA binding through the N-terminal domain. Two amino acid residues namely, Arg134 and Phe125 were shown to be the important residues involved in the RNA-binding (Gomis-Rüth et al., 2003). Mutation on these residues did affect the RNA-binding process (Thomas Hoenen et al., 2005). Bornholdt et al. (2013) stated that VP40-RNA binding is an important process to facilitate the transcription of the viral gene in the early infection stage. These studies implied that VP40 plays a vital role in Ebola virus life cycle, thus this protein will be a good target for the drug development to treat the disease.

1.2 PROBLEM STATEMENT

To date, there is no FDA-licensed drug is currently available to solve EVD and most of the potential drugs to treat EVD are yet to be used in human clinical trials (Yuan, 2015). Currently, a trial vaccine called rVSV-ZEBOV is being used to control the Ebola virus outbreak in DRC as the vaccine showed effectiveness in protecting people from EVD during a trial in Guinea in 2015 (Henao-Restrepo et al., 2015). Besides that, many virtual screening studies has been carried out to identify potential lead ligands to block the RNA-binding site on VP40 (Abazari et al., 2015; Karthick et al., 2016; M Alam El-Din et al., 2016; Mirza & Ikram, 2016; Raj & Varadwaj, 2016;

Shah et al., 2015), however, none of the potential ligands has been tested or validated in experiment study. Therefore, there is room for improvement in the optimisation of the potential ligands through computational approach for better binding at the inhibition site. Computer-aided drug design has become a valuable approach in identifying potential compounds for the drug development (Katsila et al., 2016). For example, in the quest of identifying a small-molecule inhibitor for the TGFβ-1 receptor kinase, two different research groups used different approach where one at the Eli Lilly used conventional lab approach while another group at Biogen Idec applied the *in silico* method. The *in silico* approach turned out to produce similar result of potential compound as being found experimentally in the lab (Sawyer et al., 2003; Shekhar, 2008; Singh et al., 2003). Besides that, several drugs have been successfully discovered and optimised using computational approach such as Captopril, Dorzolamide and Saquinavir (Talele, Khedkar, & Rigby, 2010). Therefore, this study focused on improving existing ligands' binding at the VP40-RNA binding site via molecular docking and molecular dynamics simulations approach and to assess its properties for developing it further as Ebola VP40-RNA binding site inhibitor.

1.3 OBJECTIVES

This study was carried out to help in finding potential Ebola inhibitors by targeting the VP40-RNA binding site. Three objectives were outlined for this study which are:

- To identify potential ligands that can inhibit the VP40-RNA binding site.
- To predict the ADME (Absorption, distribution, metabolism, excretion) properties of the potential ligands.

- To investigate ligands' binding affinity at the binding sites through molecular dynamics (MD) simulation.

1.4 HYPOTHESES

To achieve the study's objectives, the research was conducted based on several hypotheses. Potential new ligands that can block the VP40-RNA binding site can be identified and optimised based on the chemical structure or scaffold of the previously studied ligands through similarity search in database and virtual screening approach. Meanwhile, the ADME properties of the potential ligands can be predicted using SwissADME, a free web tool to evaluate pharmacokinetics and drug-likeness of small molecules. Binding affinity of the ligands at the binding site can be further monitored using MD simulation tools by analysing the RMSD, RMSF, ligand clustering and binding free energy.

1.5 SIGNIFICANCE OF THE STUDY

Ever since the 2014-2016 outbreak, the calling for further research on understanding EBOV life cycle is more urgent than before. Even though the outbreaks were mainly occurred in West Africa, current mobility in term of transportation might spread the disease further from its origin (Gonzalez, Wauquier, & Vincent, 2018). The situation can be worsened in the future due to the increase in human population which eases the transmission of disease through human to human transmission. Several imported cases had occurred in the past to several countries like Russia, Nigeria, Senegal and Spain due to the travelling from the outbreak origin (Weyer et al., 2015). Since there is no approved drug or vaccine to treat the diseases till recent time, contributions on

understanding Ebola virus as well as developing potential drugs for the disease are highly pressing for safety measure purposes in future.

CHAPTER TWO

LITERATURE REVIEW

2.1 EBOLA VIRUS DISEASE AND ITS PANDEMIC

Ebola virus disease or EVD has existed for more than 40 years concentrated in the African countries. EVD is caused by Ebola virus, which is a filamentous virus belonging to Filoviridae family together with Marburg virus and Cueva virus (Kuhn et al., 2012). To date, five different species of Ebola virus have been identified namely Zaire ebolavirus (EBOV), Sudan ebolavirus (SUDV), Taï Forest ebolavirus (TAFV), Reston ebolavirus (RESTV) and Bundibugyo ebolavirus (BDBV). As reviewed by Weyer et al. (2015), the highest fatality rate of the disease has been associated with EBOV (90% fatality rate) followed by SUDV with the average fatality rate of 50%. BDBV is the most recent Ebola virus species that was firstly emerged in 2007 in Uganda with an average fatality rate of 30%. RESTV is the only species of the virus that was discovered outside African continents and is considered to be non-pathogenic to human. Meanwhile, the TAFV which formerly known as Côte-d'Ivoire ebolavirus has caused only one non-fatal infection in human in 1994 but very deadly in chimpanzees. Among the Ebola virus species, EBOV is the most studied species due to its high fatality rate and occurrence.

EVD, formerly known as Ebola haemorrhagic fever was firstly appeared in year 1976-1979 period where the first outbreak occurred in Sudan, near the border with the Democratic Republic of Congo (DRC) by SUDV between June and November 1976. That particular outbreak killed 150 out of 284 victims (53% fatality rate) (WHO, 1976). The second outbreak occurred in DRC between August and

November 1976 that caused 284 deaths out of 318 victims (89% fatality rate) (Breman et al., 1978). The third outbreak emerged in 1979 between July and October in Sudan with the fatality rate of 65% (Heymann et al., 1980). After 15 years of silence, EVD re-emerged in Taï National Park, Ivory Coast in 1994 that led to the discovery of TAFV that caused massive deaths among chimpanzees (Formenty et al., 1999). Since then, EVD cases have been reported almost annually leading up to its biggest outbreak in 2014 in West Africa (Weyer et al., 2015). Prior to the 2014 outbreak, EVD outbreaks occurred randomly and rarely appeared in the same location where most affected countries were in the Central Africa namely Congo, DRC, Sudan, Gabon and Uganda (Figure 2.1).

Ebola outbreak in 2014 was caused by Zaire species that affected three countries; Guinea, Liberia and Sierra Leone. In March 2014, World Health Organization (WHO) declared the emergence of the outbreak in Guinea which could be the biggest and fatal outbreak in Ebola history (WHO, 2014). Throughout two years, 28,646 EVD cases were reported in the three countries, with 11,323 deaths. Besides that, several travel-associated cases had been reported during that time in Nigeria, USA and Spain but only with small casualty (WHO, 2016). The 2014-2016 Ebola outbreaks occurred in major urban areas while previous Ebola outbreaks were occurred in rural and secluded areas. In August 2018, an EVD outbreak was declared by the Ministry of Health of the Democratic Republic of the Congo that took place in North Kivu Province. By March 2019, almost 1000 cases were confirmed with 621 deaths (WHO, 2019).

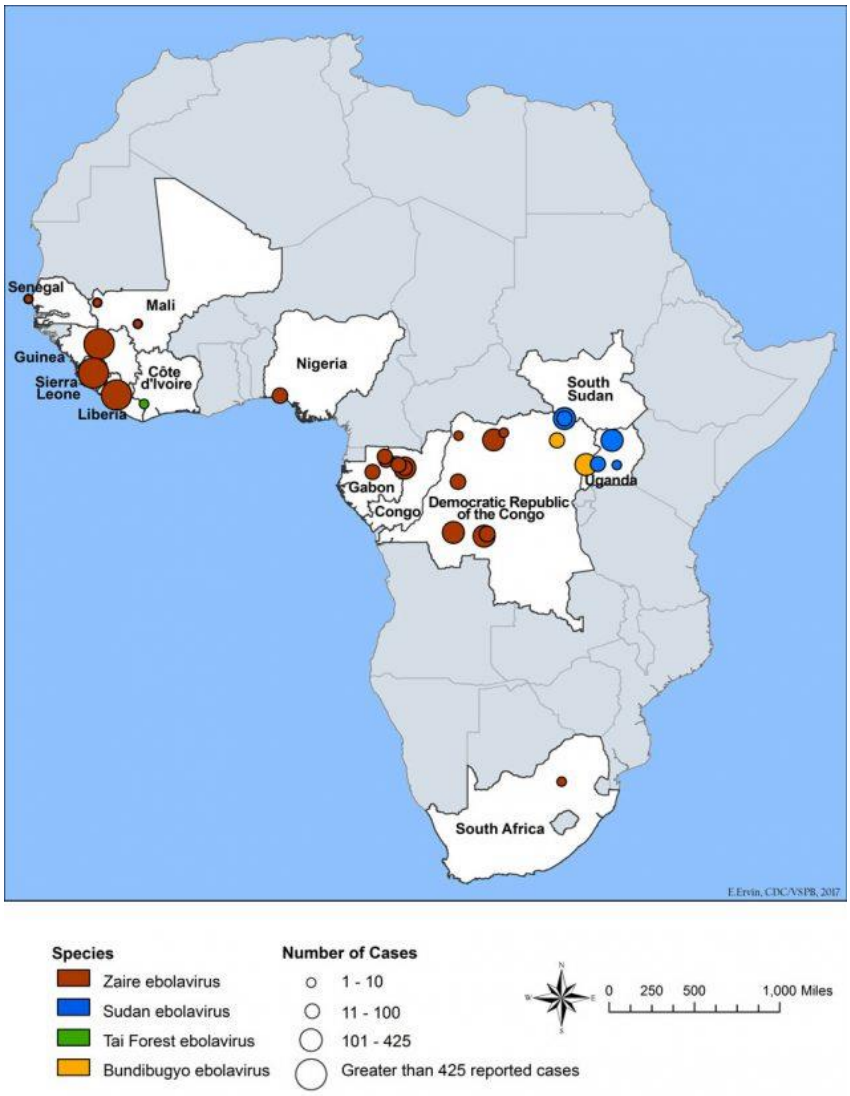

Figure 2.1 Distribution of the Ebola Virus in Africa from Its Discovery in 1976 till the Largest Outbreak That Ended in 2016. Most of the Outbreaks Were Associated with Zaire and Sudan Species.
Adapted from https://www.cdc.gov/vhf/ebola/history/distribution-map.html

2.2 EVD TREATMENT AND THERAPY

Currently, there are no approved treatments or medications for EVD around the world. Current treatment of EVD includes the administration of supportive care and treatment strategies like rehydration with intravenous or oral fluids (Geenen et al., 2014). However, due to the severity of the 2014-2016 Ebola outbreak, WHO has allowed several experimental medicines and vaccines to be used to combat the