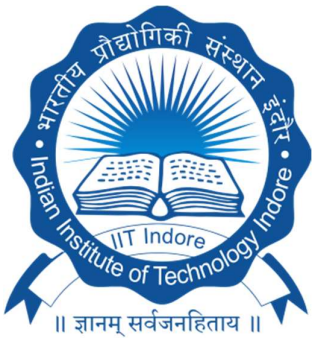
Viral genome targeting for anti-viral drug discovery and vaccine development against Monkeypox virus

M.Sc. Thesis

By
NIKHIL KUMAR



DEPARTMENT OF BIOSCIENCES AND BIOMEDICAL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY INDORE

MAY 2023

Viral genome targeting for anti-viral drug discovery and vaccine development against Monkeypox virus

A THESIS

Submitted in partial fulfillment of the
requirements for the award of the degree
of
Master of Science

by
NIKHIL KUMAR

DEPARTMENT OF BIOSCIENCES AND BIOMEDICAL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY INDORE

MAY 2023

INDIAN INSTITUTE OF TECHNOLOGY INDORE

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the thesis entitled **VIRAL GENOME TARGETING FOR ANTI-VIRAL DRUG DISCOVERY AND VACCINE DEVELOPMENT AGAINST MONKEYPOX VIRUS** in the partial fulfillment of the requirements for the award of the degree of **MASTER OF SCIENCE** and submitted in the **DEPARTMENT OF BIOSCIENCES AND BIOMEDICAL ENGINEERING, Indian Institute of Technology Indore**, is an authentic record of my own work carried out during the time period from August 2021 to May 2023 under the supervision of *Dr. Amit Kumar* (Professor, Discipline of Biosciences & Biomedical Engineering).

The matter presented in this thesis has not been submitted by me for the award of any other degree of this or any other institute.

Nikhil Kumar
09/05/2023

Signature of the student with date
(NIKHIL KUMAR)

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

Signature of the Supervisor

(Prof. Amit Kumar)

NIKHIL KUMAR has successfully given his/her M.Sc. Oral Examination held on **May 9, 2023**

Signature(s) of Supervisor(s) of MSc thesis
Date:

P.V. Kodgire

Convener, DPGC
(Prof. Prashant Kodgire)
Date: 09/05/2023

Parimalkar

Signature of PSPC Member
(Dr. Parimal Kar)
Date: 09/05/2023

Signature of PSPC Member
(Dr. Abhijeet Joshi)
Date: 09/05/2023

ACKNOWLEDGEMENTS

"Arise! Awake! and stop not until the goal is reached."

Swami Vivekananda

I would like to express my sincere gratitude to my advisor, *Professor Amit Kumar*, for their invaluable guidance and support throughout my master's program. Their expertise and encouragement helped me to complete this research and write this thesis.

I also place on record my sense of gratitude to the members of the PSPC committee, *Dr. Parimal Kar* and *Dr. Abhijeet Joshi* for their insightful comments and for sharing their tremendous experience with me. I am thankful to *Professor Amit Kumar* (Head, Discipline of Biosciences & Biomedical Engineering) for his continuous support in various aspects. I wish to express my sincere thanks to all faculty members of BSBE discipline for all their suggestions and support. I would like to acknowledge all the teachers I learned from since my childhood; I would not have been here without their guidance, blessings, and support. I am also thankful to all non-teaching staff of the BSBE family, Mr. Arif Patel, Mr. Murphy Bhaskar Ganveer, and Mr. Gaurav Singh. Their insights and guidance were instrumental in helping me to shape my research and write this thesis.

I thank my lab seniors, Krishna Singh, Ashmad Nayak, Aritra Chakraborty, Pronamika Chetia, Aakriti Singh, and Priya Gupta for their constant support and motivation. I am lucky to have a colleague like Rishav Nag, have a great scientific temperament and is fun to work with him. I have learned a lot from them in my MSc journey. Thank you for tolerating me.

I am Lastly, I would like to thank my family for all their care, love, affection, and encouragement. I sincerely express my love and gratitude to my father Mr. Sunil kumar for his unconditional

love and trust and deeply thankful to my friends for their love and support during this process. Without their encouragement and motivation, I would not have been able to complete this journey.

Finally, I would like to extend my sincere gratitude to all of the participants in my study. Their willingness to share their experiences and insights has been invaluable to my research and has helped to make this thesis a success. Thank you for your time and contribution.

I would also like to express my sincere thanks to IIT Indore for its infrastructure and the Department of Biotechnology (DBT), India, for my fellowship. I am grateful to everyone who has supported me throughout this process. Without your help and guidance, this thesis would not have been possible.

(Nikhil Kumar)

Dedicated

to

My

Family

Abstract

G-quadruplexes (GQs) are one of the most extensively researched non-canonical secondary structures of nucleic acids and are ubiquitously present in all species, from viruses to prokaryotes and eukaryotes. GQs are formed by staking of more than one G-tetrad, and each G-tetrad is made by non-canonical reciprocal interaction of Guanine residues arrayed in a plane. G-quadruplexes are present in the regulatory regions, mostly in promoter, intron, and telomeres. Due to the presence of these motifs in regulatory regions, they assist in various biological processes, for example, replication, transcription, repair & recombination, and telomere maintenance. The development of novel therapeutics relies on identifying novel drug targets that are evolutionary conserved in monkeypox virus and their nucleic acid act as better receptors for developing novel anti-viral drugs. Multiple studies have been done on several viruses including Ebola, Zika, Human adenovirus, HIV, and HCV etc, showing the presence and regulatory role of GQ motifs in viral replication and amplification. Therefore, they are used as a potential anti-viral therapeutic target. Similarly, we move forward with the monkeypox virus, In our initial studies we identified the highly conserved GQ motifs in the genome of monkeypox virus that might play a critical role in their growth and amplification. We further analyzed with G-quadruplex ligands like Braco-19 and analyze their interaction, binding, and stability in the presence of these drugs. All these studies provide a glimpse of the regulatory role of the G-quadruplex motif in the monkeypox virus which can lead to novel therapeutic approaches and drugs.

Targeting another important cause of the Monkeypox virus outbreak, not a single specific vaccine is in the market against it. The development of vaccines can help to control the spread of the virus and prevent new cases from occurring. Our study provides a brief immunoinformatics approach to finding a novel vaccine candidate against Monkeypox virus infection. Here we use different approaches such as the selection of different epitopes, MD simulation, and in silico cloning analysis helped in the generation of an effective vector that is used for the multi-epitope vaccine expression. Our comprehensive in silico research offers a foundation for the creation of a Monkeypox virus vaccine as a prevention measure.

LIST OF PUBLICATIONS

- ***Kumar N[1]****, Nayak A[1], Chakraborty A[1], Kumar A** (2023) *In-silico* designing a novel TLR4-mediating multiepitope vaccine against monkeypox *via* advanced immunoinformatics and bioinformatics approaches. *J Biomol Struct Dyn*. (Communicated)
- ***Kumar N[1]****, Nayak A[1], Kumar A** (2023) Identification and characterization of conserved G-quadruplex forming motifs in the Monkeypox virus genome and their interaction with G-quadruplex specific ligands. *Sci Rep.* (Communicated)
- ***Kumar N[1]****, Nayak A[1], Kumar A*.* Biological Relevance of RNA G-quadruplexes in the human long non-coding RNA (Manuscript under preparation)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

NOMENCLATURE

λ	Wavelength
ε	Extinction coefficient
π	Pi
Å	Angstrom
°C	Degree centigrade
°	Degree
μL	Microliter
mL	Milliliter
L	Liter
nm	Nanometer
pmol	Picomole
nmol	Nanomole
nM	Nanomolar
μM	Micromolar
mM	Millimolar
M	Molar
S	Second
mins	Minutes
h	Hour

ACRONYMS

A	Adenine
C	Cytosine
CD	Circular dichroism
cDNA	Complimentary DNA
D/N	Drug/nucleotide or drug/DNA
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
EMSA	Electrophoretic mobility shift assay
G	Guanine
GQ	G-quadruplex
GQs	G-quadruplex structure
G4IPDB	G-quadruplex (G4) Interacting Protein Database
HIV	Human immunodeficiency virus
ITC	Isothermal titration calorimetry
MD	Molecular dynamics simulations
MM-PBSA	Molecular mechanics/Poisson-Boltzmann Surface Area
MPXV	Monkeypox virus
NMR	Nuclear magnetic resonance
ORF	Open reading frame
PCR	Polymerase chain reaction

Chapter 1

Introduction

1.1. Monkeypox Virus

Monkeypox virus (MPXV) belongs to the Poxviridae family of genus orthopoxvirus, which includes other notable members such as smallpox virus and cowpox virus. The virus was first identified in monkeys, but it may also infect humans and cause monkeypox, a rare but potentially dangerous disease. Although outbreaks have happened elsewhere in the world, Central and West Africa are where the monkeypox virus is most commonly found. It possesses a double-stranded DNA genome with a lipoprotein envelope and a size range between 200-250 nm, causing diseases in animals and humans when come in contact with infection (zoonotic diseases). MPXV completes its lifecycle in the cytoplasm instead of the nucleus of infected cells [1,2].

1.1.1. Structure and Classification of Monkeypox Virus

The monkeypox virus is a huge, intricate, multilayered virus. The virus has an ovoid or brick-shaped form, measuring roughly 200–250 nm in length and 140–170 nm in breadth. The envelope, which is composed of lipids and viral proteins, is the virus' outermost coating. The tegument, a layer of proteins under the envelope that plays a role in the growth and development of the virus, is present. The genetic material, which is a double-stranded DNA molecule that codes for the virus' proteins and enzymes, is found in the core of the virus [2].

The Orthopoxvirus genus of the Poxviridae family is where the monkeypox virus is categorized. The variola virus, which causes smallpox, the vaccinia virus, which is used in smallpox vaccinations, and the cowpox virus are all members of the orthopoxvirus genus and are closely related to the monkeypox virus. The enormous, intricate architecture of orthopoxviruses and their capacity

to proliferate in the cytoplasm of host cells are characteristics of these pathogens.

Monkeypox virus is further divided into two clades, or subtypes, within the Orthopoxvirus genus based on chromosomal and immunological variations. The Congo Basin Clade and the West African Clade are the names of the two clades. While the West African clade typically causes a milder form of the disease, the Congo Basin clade is linked to more acute cases of monkeypox [3].

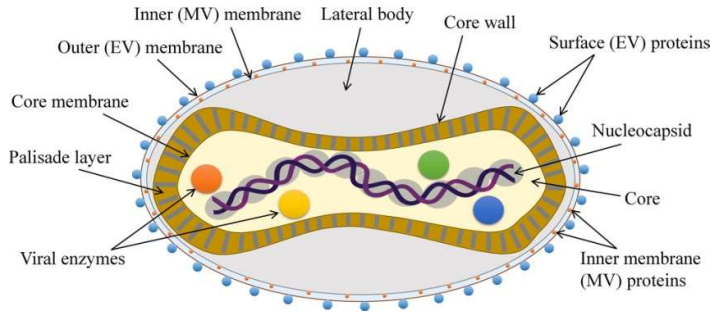


Figure 1.1. Structure of a Monkeypox virus **[2].**

1.1.2. Viral infection

Viruses infect hosts by attaching to specific receptor sites on the surface of host cells, which allows them to enter the cell. Once inside the host cell, the virus uses the host's cellular machinery to replicate and produce new virus particles. This process can cause damage to the host cell, leading to the symptoms of infection.

Monkeypox virus is transmitted to humans through contact with infected animals, such as rodents or primates, or through contact with infected humans. The virus causes a range of symptoms, including fever, headache, muscle aches,

and a rash that usually begins on the face and then spreads to other parts of the body. In severe cases, monkeypox can cause complications such as pneumonia, encephalitis, and even death [3].

1.2. Monkeypox virus outbreak – A global challenge.

In recent years, monkeypox outbreaks have been reported in several African countries, including Nigeria, the Democratic Republic of Congo, and Cameroon. In addition, cases of monkeypox have been imported into other countries, such as the United States and the United Kingdom. The most recent outbreak of monkeypox occurred in Nigeria in 2021, where over 200 cases were reported. This outbreak was particularly concerning because it involved the first reported case of human-to-human transmission of monkeypox virus in Nigeria, and it also appeared to be associated with a higher mortality rate than previous outbreaks.

Another recent outbreak of monkeypox virus occurred in the United Kingdom in 2018, with three confirmed cases reported in individuals who had recently traveled from Nigeria. The individuals were hospitalized with symptoms of monkeypox virus infection, but all made a full recovery [2].

1.2.1. Causes of Monkeypox outbreak

Monkeypox virus outbreaks are caused by the transmission of the virus from infected animals to humans, or from human-to-human transmission. The virus is believed to circulate among certain species of animals, such as rodents and primates, in Central and West Africa where the virus is endemic. Humans can become infected through direct contact with infected animals, such as handling their blood, bodily fluids, or meat, or through contact with contaminated objects, such as bedding or clothing.

Overall, the causes of monkeypox virus outbreaks are multifactorial, involving a complex interplay between the virus, animal reservoirs, human behavior and contact with wildlife, and social and environmental factors. Efforts to prevent

and control outbreaks require a multi-disciplinary approach that involves public health officials, veterinarians, environmentalists, and communities at risk for infection [1].

1.2.2. Spread of Monkeypox virus

Monkeypox virus is primarily transmitted to humans through contact with infected animals, particularly rodents such as Gambian giant rats and rope squirrels, or through contact with infected primates, such as monkeys and chimpanzees. The virus can also be transmitted from human to human through respiratory droplets, contact with bodily fluids, or contact with contaminated objects, such as bedding or clothing.

1.2.3 Diagnosis and possible treatment for Monkeypox virus

Diagnosing monkeypox can be challenging due to its similarity with other viral illnesses, such as chickenpox and smallpox. A diagnosis of monkeypox is usually confirmed through laboratory testing of blood, skin lesions, or other bodily fluids. The most commonly used tests include polymerase chain reaction (PCR) to detect the virus's genetic material, virus isolation, and serological testing to detect the presence of antibodies against the virus.

Treatment may include antiviral drugs such as cidofovir or brincidofovir, which have shown some effectiveness against the virus in animal studies and in some human cases. Vaccination with the smallpox vaccine has been shown to be effective in preventing monkeypox, as the two viruses are closely related. In conclusion, there is no specific treatment or vaccine for monkeypox, but supportive care can help manage symptoms and complications [4,5].

1.2.4 Possible solutions

The development of novel drugs to combat the Monkeypox virus is the need of the hour. Still, the discovery and development of specific antiviral drugs or

vaccines are in a declining phase due to failures at multiple ends. One of the significant reasons is an inadequate selection of molecular targets to a lack of novel strategies. Another approach includes identifying host factors critical to virus-host interaction and blocking it. In the current scenario, identifying the novel and highly conserved molecular target is a crucial step in drug discovery and vaccine development.

1.3. Therapeutic potential of G-quadruplexes in Viruses

A virion is an infectious particle made up of a capsid and a protein shell that contains the genetic material of the virus, which can be either single-stranded DNA or RNA. These viruses are strong intruders because they take advantage of the host's machinery to grow and multiply [8]. G-quadruplexes are commonly described in both people and viruses, where they play regulatory functions in viral invasion and the conclusion of the viral life cycle [6,7]. The first virus with GQ structures was identified as Human Immunodeficiency Virus-1 (HIV-1). Three putative conserved GQ motifs were identified by computational analysis in the nef gene, and the expression of the gene was significantly reduced when the host-infected cells were treated with the GQ-specific ligand, TMPyP4. GQ motifs have a critical role in the expansion and pathogenicity of HIV, making them potential targets for antiviral therapy and opening up new avenues for its development [9].

The B-lymphocyte gets invaded by the Epstein-Barr virus (EBV), which is then kept alive as an episome by host genome inclusion. The GQ-specific small chemical BRACO-19 interferes with this RNA-protein interaction and diminishes viral multiplication in host cells [10]. The conserved GQs in the preS2/S gene's promoter region of the hepatitis B virus (HBV), which are necessary for viral assembly, have been found to increase the gene's transcription [11]. Four GQ motifs were discovered in ZIKV's NS2, NS4B, and NS5 nonstructural proteins, according to another study. The two GQ-specific ligands, TMPyP4 and BRACO-19, significantly inhibited viral multiplication and protein synthesis [12]. The L gene of the Ebola virus, which codes for RNA-dependent RNA polymerase (RdRp), was recently found to include a GQ-

forming motif [13]. These structures may offer a novel antiviral therapeutic approach due to their existence in the Ebola virus genome. The G and L genes of the Nipah virus, which is a member of the Paramyxoviridae family, contain two highly conserved GQs that are used to synthesize the glycoprotein (G) and the RNA-dependent RNA polymerase protein, respectively. For viral entrance and replication within the host cell, these genes are necessary [14]. In conclusion, these findings provide alternate therapeutic approaches to focus on the G-quadruplex structures in the viral RNA genome. In Table 1.3, a collection of G-quadruplex motifs discovered in viral genomes is provided.

1.4. Aim of the thesis

The aim of the current MS.c. thesis work is focused on the identification, characterization, and validation of several novel conserved G-quadruplex motifs as a promising anti-viral drug target in the Monkeypox virus. As G-quadruplex are present in the regulatory regions, mostly in a promoter, intron, and telomeres. Due to the presence of these motifs in regulatory regions, they assist in various biological processes for example replication, transcription, repair & recombination, and telomere maintenance. Multiple studies have been done on several viruses including Ebola, Zika, Human adenovirus, HIV, and HCV, which possess the GQ structure in the viral genome. These structures have been targeted by GQ-specific small molecules leading to their use as antiviral therapy. Therefore, we aim to identify the role of these G-quadruplex motifs in the *Monkeypox virus* genome and these evolutionarily conserved structures can be exploited as therapeutic drug targets. In summary, the current thesis aims to provide comprehensive information about the presence of conserved GQ structures in the *Monkeypox virus* genome and provides a platform to exploit these unique structures as attractive anti-bacterial therapeutic targets.

Targeting another important cause of the *Monkeypox virus* outbreak, not a specific vaccination against it. A screening was initiated for the selection of different epitopes and added with linkers and adjuvants to increase the immunogenicity and antigenicity of the vaccine. Refinement and analysis of

vaccine construct using virtual screening, molecular docking, MM-PBSA, and Molecular dynamics approach for safe and effective vaccine development.

1.5. Scope of the thesis

In recent years, an outbreak of viruses like corona, monkeypox, etc. causes havoc and a pandemic in world society. This requires scientific intervention for anti-viral drug discovery and vaccine development. The present thesis works revolve around tackling one of the potential pandemic viruses, the Monkeypox *virus* by targeting antiviral drug discovery and effective vaccine development. First, for antiviral drug discovery, the work embodies the identification and prominence of evolutionary conserved G-quadruplex motifs in the *Monkeypox virus* genome. Ligands or small molecules having high specificity for the viral GQ structures over non-specific targets can emerge as promising anti-viral therapeutics, giving hope for the prevention of the potential viral pandemic.

Second, the thesis delves into the development of a vaccine for the Monkeypox virus would be particularly beneficial in areas where the Monkeypox virus is endemic, such as Central and West Africa, and can help to limit the spread of the virus and prevent new cases from occurring. Overall, the development of a vaccine for the Monkeypox virus is an important step in protecting public health and reducing the impact of this disease.

In summary, the scientific findings described in the thesis aim to put forward a step closer to the search for novel strategies for anti-viral drug discovery and vaccine development against the Moneypox virus.

Chapter 2

Material, methods, and instrumentation

2.1 Materials

All the DNA oligonucleotides which are used in the biophysical characterization or taken as positive controls or negative controls like *cmyc*-GQ, *Bcl2*-GQ, etc. were procured from Sigma-Aldrich Chemicals Ltd. (St. Louis, MO, USA) and/or Integrated DNA Technologies (Iowa, United States). The list of oligonucleotides used is enlisted in the respective chapters.

Table 2.1. The sequences for G-quadruplex motifs used for this study are listed below:

S.no	Name	Sequence (5'- 3')
1.	B3R	TAGGAAAAGGTGGATTCGGTA
2.	DUTPase	GAGGAGATCAAGGGTTTGGATCAACAGGAC
3.	B12R	AAGGAGGAGAACATGGATAAGGTT
4.	A25R	ACGGATGGGATTCTTCGGTAACTTGGAG
5.	I4L	ATGGCTATTCTGGCGGCTAGAATGGCA
6.	L6R	ATGGTAGACTAGGTGCTATGGATGGGGCA
7.	F4L	TAGGTGATGCTGGTTCTGGAGATTCTGGAG
8.	EEVMaturation Protein	ACGGTGCTAGGGTTATTTGGATGGAT
9.	ESR	ATGGTGTATTGGACCTGGTAGACGGAA
10.	IMV Membrane Protein	CAGGTTGTTGGGGTTGGTATTGGTT
11.	A19R	CAGGTAGAGGCAGTCGGATTGGAA
12.	A49R	ACGGTTACTGGACCAGTGGGGCAACTGTGGAT
13.	c-myc	TGAGGGTGGTGAGGGTGGGGAAGG

The chemical reagents which are used in buffer and gel preparation are NaCl, KCl, LiCl, MgCl2, K2HPO4, KH2PO4, NaH2PO4, Na2HPO4, NaOH, HCl, 3-(Trimethylsilyl) propionic-2, 2, 3, 3-d4 acid sodium salt (TSP), 3-(Trimethylsilyl)propionic-2,2,3,3-d4 acid sodium salt (DSS), D2O, Tris base, Boric acid, Ethylenediaminetetraacetic acid (EDTA), DMSO, Ammonium Persulfate (APS), Tetramethylethylenediamine (TEMED), Acrylamide, Bis-Acrylamide, ethanol, isopropanol, methanol, urea, etc. were purchased from Sigma-Aldrich Chemicals Ltd. (St. Louis, MO, USA) and were mostly of HPLC purified or molecular grade (wherever applicable). The GQ-specific ligand molecules such as BRACO-19 were also procured from Sigma-Aldrich Chemicals Ltd. (St. Louis, MO, USA).

Other reagents for media preparation like Agarose, Luria broth, Luria agar, and antibiotics like Ampicillin, Kanamycin, Chloramphenicol, *Taq polymerase,* etc. were bought from Himedia Laboratories, India. The typical plastic wares used were from Tarsons Products Pvt. Ltd. (Kolkata, India), and Sartorius (Sartorius Corporate, Germany) water purification system was used to obtain molecular grade water for preparation of media and buffer.

2.2. Sample Preparation

2.2.1. Oligonucleotide preparations

According to the manufacturer's instructions, Mili-Q water was used to dissolve the oligonucleotides, and absorption was measured at 260 nm. Then, by applying Beer’s-lambert law its concentration was calculated.

2.2.2 Sample preparation for 1D proton NMR spectral analysis

DNA samples were dissolved in 10x phosphate buffer, pH 7.2 containing 100 mM KCl, in H2O: D2O ratio of 9:1. To rectify the problem of paramagnetic defect which causes broadening of peak during measurement, the addition of EDTA 0.1 mM can nullify the effect. For setting the zero of the experiment DSS (1μM) was taken as a reference. Mix it with a pipette properly and add 200μl sample in Shigemi tubes for an experiment.

2.2.3 Sample preparation for Circular Dichroism

In independent tests, each Circular Dichroism (CD) spectrum was captured at a final concentration of 20 µM of oligonucleotides in Tris buffer (10 mM, pH 7.4) that contained 50 mM concentrations of K+, Na+, Mg2+, and Li+ for each of four distinct cations. Before the experiment, The samples were heated at 92 ºC for 10 minutes and after that, it is cooled.

2.2.4 Sample preparation for electrophoretic mobility shift assay

A conventional G-quadruplex and a linear duplex DNA oligonucleotide of the same number of bases (Bcl-2 DNA and c-myc DNA) were used as the respective negative and positive controls. For each experiment, a 20% native polyacrylamide gel (PAGE) was created in a 1X Tris-Borate-EDTA (TBE) buffer. In the appropriate buffers containing K+, Na+, Mg2+, and Li+ cations (100 mM), 20 µM of each oligonucleotide was produced, and it was then annealed by heating at 90 °C for 10 minutes, followed by gentle cooling at room temperature. Before being loaded onto the gel every sample was blended with 6X orange dye.

2.2.5. Sample preparation for isothermal titration calorimetry (ITC)

Before each ITC experiment, DNA samples were dissolved in 1x phosphate buffer containing 50 mM KCl, and then they were heated to 90 ºC for 10 minutes, annealed, and then slowly cooled to room temperature. Small compounds (BRACO-19 etc.) were produced as a stock and working solution in the same buffer condition and both DNA and small compounds were degassed before the experiment.

2.3. Methodology and Instrument Specifications

2.3.1. Virus genome retrieval

The whole genome sequences of the virus under study were downloaded from the NCBI Genome (https://www.ncbi.nlm.nih.gov/genome) [15] and G-quadruplex motif analysis was conducted.

2.3.2. Bioinformatics Tools used for G-quadruplex Screening

There are various computational tools for mining the GQ motifs in genomes like Quadruplex forming G-rich sequences (QGRS)mapper[16], and Nucleic acid G-quadruplex structure (G4) Interacting Proteins database (G4IPDB)[17], etc. after mining of GQ motifs conservation analysis was done by manual clustering of sequence, performed by using NCBI Genebank coordinates.

2.3.2.1. QGRS Mapper

QGRS mapper (http://bioinformatics.ramapo.edu/QGRS/) is a web-based server that helps in accessing and estimating the GQ motifs and also helps in searching and locating gene or nucleotide sequences from the NCBI database. It uses the below-given formula to predict the G-quadruplex:

GxNy1GxNy2GxNy3Gx

Where x = a number of guanine tetrads in the G-quadruplex and y1, y2, y3 = length of gaps between G-tetrads.

2.3.2.2. G4IPDB

An in-house developed web tool G4IPDB server (http://people.iiti.ac.in/~amitk/bsbe/ipdb/update/tool.html) is also used for predicting the GQ motifs. This tool is based on cG/cC ratio, higher the ratio more the probability of the sequence having a G-quadruplex structure.

2.3.3. Isothermal titration calorimetry assay (ITC)

Isothermal titration calorimetry (ITC) is the most used technique in biophysical studies, based on detecting the change in heat energy during experiments to estimate the binding interaction between two molecules. By evaluating different thermodynamic parameters such as association constant (Ka), enthalpy (ΔH), and entropy (ΔS) using the following equation.

$\Delta G = -RT \ln Ka = \Delta H - T\Delta S$ (2)

Where R is the gas constant and T is the temperature.

The ITC contains two cells, one for the reference (water) and another one is a sample cell. The starting temperature difference between two cells sets as baseline temperature. A small amount of reactant (Braco-19) is injected into sample cell, as they reaction occurs temperature rises and the difference

between sample and reference cell temperature is recorded. After each reaction or injection, calorimeter comes to baseline temperature. Reaching the equilibrium, where reaction temperature is similar to baseline temperature. Recorded data as plotted as observed heat at each reaction and the time of injectio ns (Figure 2.1)

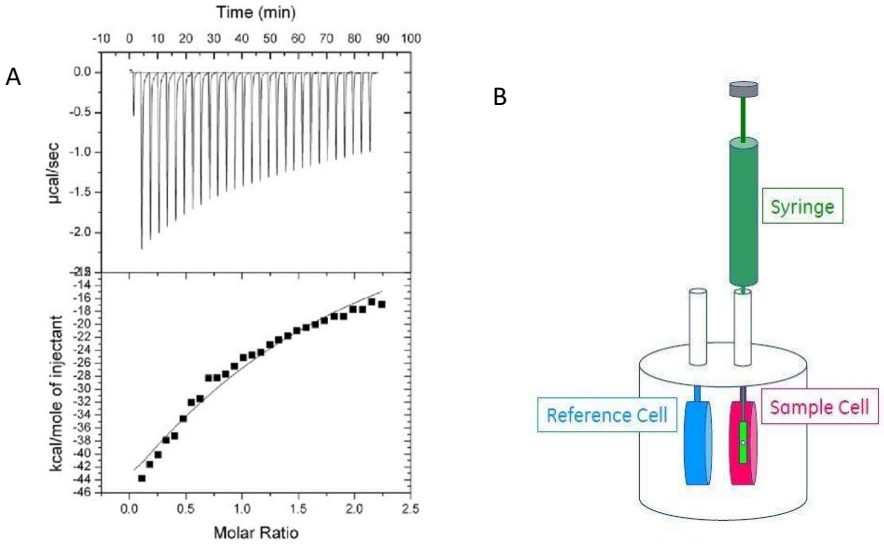


Figure 2.1: A) Schematic diagram of the ITC instrument and B) example of data plotted of the binding isotherm.

2.3.4. Circular Dichroism (CD) spectroscopy

CD spectroscopy technique is used to determine the secondary structure and also estimate interaction patterns with ligand molecules or proteins. It is based on difference between absorption of left and right circularly polarized light that can be formulated as[18].

$CD = Al - Ar$ (3)

Where *Ar* and *Al* stand for the absorptions of right and left circularly polarized light respectively.

The Jasco J-185 Spectropolarimeter (Jasco Hachioji, Tokyo, Japan) was used in the current investigation to carry out CD studies for each oligonucleotide in various buffer conditions at 25 ºC with a scanning rate of 20 nm/min from 220 to 320 nm. There was a Peltier Junction temperature controller in the instrument.

In four consecutive studies, spectra were taken in a cuvette with a 1 mm path length and a final concentration of 20 µM of all oligonucleotides in Tris buffer (10 mM, pH 7.4) with 50 mM concentrations of K+, Na+, Mg2+, and Li+. Before every measurement, a blank spectrum was recorded and subtracted from the CD spectrum of the sample prepared in the appropriate buffer to prevent signal contribution from the buffer (Figure 2.2).

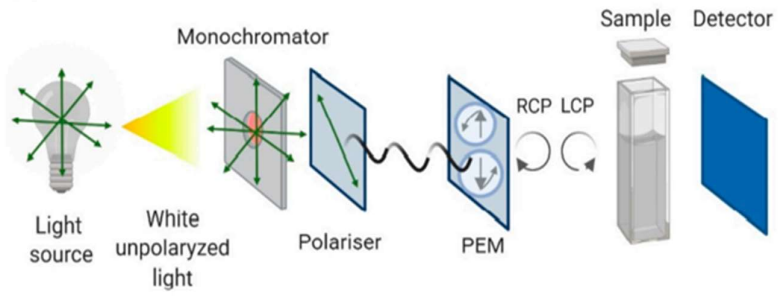


Figure 2.2: Schematic illustration of the CD instrument. PEM is a photoelastic modulator that produces right circularly polarized light (RCP) and left circularly polarized light (LCP), and PMT is a photomultiplier and works as a detector.

2.3.5 Nuclear magnetic resonance (NMR) spectroscopy

The nuclear magnetic resonance (NMR) spectroscopy technique is a primary quantitative method in biophysical analysis. It helps in structural analysis of a molecule such as the Watson crick base (Nucleic acid) pairing gives a chemical shift signal at 12 to 14 ppm whereas the Hoogsteen base pairing present in the G-quadruplex structure give a chemical shift signal between 10 to 12 ppm[19].

A radiofrequency generating coil is positioned between two magnetic poles in a fundamentally simple classical 60-MHz NMR spectrometer. The nucleus absorbs electromagnetic energy and alters the proton's spin orientation. A detection coil is positioned parallel to the RF oscillator coil. It causes a radiofrequency signal to be induced in the detector coil at the time that nuclear

spins are reoriented. This is noted by the instrument as an NMR peak or resonance signal.

In the current thesis, NMR spectroscopic investigation was carried out using an AVANCE 400/500 MHz BioSpin International AG, Switzerland device outfitted with a 5 mm broadband inverse probe. Using a Broad Band Inverse probe (BBI probe), all 1D 1H NMR spectra were captured at 298K. The Topspin 1.3 program (academic license) was used for NMR data processing, integration, and analysis.

2.3.6 Electrophoretic mobility shift assay (EMSA)

The well-known and widely-used method to distinguish between the size and molecularity of biological macromolecules is EMSA [20]. It is based on molecules' movement in the agarose or polyacrylamide gel which depends upon its charge, shape, and size[20]. There are two different G-quadruplex conformations that could exist: intramolecular and intermolecular. Since the intramolecular GQ structures are more compact than their linear counterparts, it has been found that they move more quickly through the electrophoretic gel. Contrarily, as shown in (Figure 2.3), the intermolecular structures, which are heavier, move more slowly and appear to be positioned above the linear nucleic acid sequence or the intramolecular structures.

Additionally, EMSA is used to identify interactions between small molecules and nucleic acids based on a delay in their motion caused by the formation of a complex between them. Because a nucleic acid's molecular weight increases when it binds to a small molecule, therefore it moves more slowly than unbound nucleic acids. At 4º C and 90 Volts, electrophoresis was carried out using a vertical gel unit system. The gels were examined on the ImageQuant LAS 4000 gel doc (GE Healthcare Biosciences Ltd., Sweden) after being stained with ethidium bromide to visualize them.

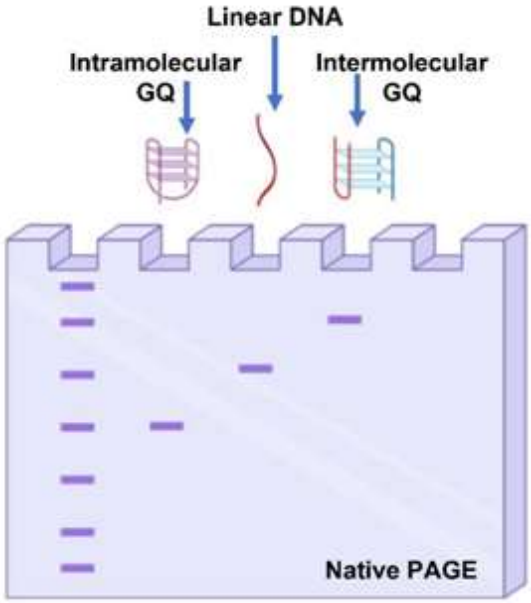


Figure 2.3: Schematic illustration of the gel mobility shift experiment. Influence of nucleic acid movement in the native PAGE gel on shape, conformation, and molecularity.

2.3.7 UV-Visible spectrophotometry

One of the most widely used methods in molecule and chemical structure analysis is UV-Visible spectrophotometry. It entails determining how much UV or visible radiation a material in solution has absorbed. Ultraviolet-Visible spectrophotometers are instruments that measure the ratio, or function of ratio, of the intensity of two beams of light in the U.V. area.

Thermal stability of a particular GQ structure with or without a ligand can also be accessed with UV-Visible spectrophotometry by tracking the changes in the reading with increasing temperature at a particular wavelength. Thus, the UV-Visible thermal denaturation assay provides us with the melting temperature (Tm) value, and higher the Tm value, higher will be the thermal stability of the secondary structure. Each sample was subjected to a UV melting study at temperatures ranging from 25 °C to 98 °C in increasing K+ concentration in buffers from 0 Mm to 200 Mm. For each of the melting studies, the heating rate was set at 1 °C/min. The ligand interaction analysis was also done, ligand was

added with various Drug/nucleic acid (D/N) ratios, and spectral and thermal melting analyses were carried out to perform ligand titrations.

2.3.8. Fluorescence Binding Assays

An electron moves from a higher energy state to a lower energy level after being stimulated to a higher energy state by consuming high-energy electromagnetic radiation. This phenomenon is known as fluorescence (Figure 2.4)[9n]. When attached to a biological macromolecule, a fluorophore undergoes conformational shifts that result in either extremely strong fluorescence or the adoption of a photo-quenched state[10-13n]. Fluorescence changes can be utilized to analyze the selectivity and binding affinity of small compounds with macromolecules. In the current study, we used this method to determine how a small molecule's fluorescence changed after it was bound to the nucleic acid. We conducted these tests using 96-well Corning half-area black plates and a Synergy H1 multi-mode microplate reader at a temperature of 25 ºC. The GQ containing DNA (20 µM) and Braco-19 (5 µM) in 10x potassium phosphate buffer containing 100mM KCl were heated at 92 °C for 10 min and then kept at room temperature for 2 hours to anneal. Fluorescence intensities were collected at the emission wavelength of 490 nm by exciting at 440 nm in the plate reader.

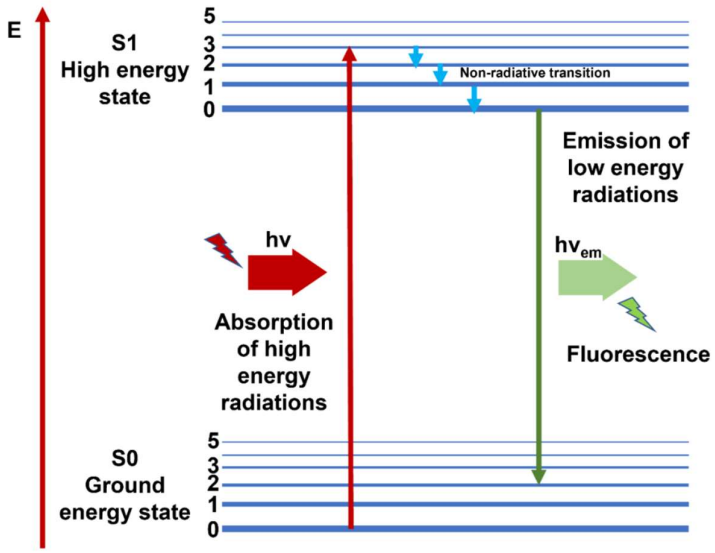


Figure 2.4: Jablonski diagram illustrating the fluorescence phenomena, in which the molecule is stimulated to a higher energy state and then releases energy.

2.3.9. Molecular Docking and Molecular Dynamics Simulations

Molecular docking is a valuable method which involves the prediction of ligand orientation and conformation within a target site. Molecular docking is a multi-step complex process with two important aims: 1) Correct prediction of activity and 2) Accurate structural modeling.

Molecular dynamic simulation is an in-silico simulation technique that informs us of the dynamic behaviour, or the physical properties of the molecules in a given system during a predetermined period. The analysis provides details about how the conformations, thermodynamic states, energy parameters, etc. of biomolecules vary and fluctuate throughout time. This platform is essential for gaining a basic understanding of the interaction profile in the fields of drug discovery and vaccine development. The programme for molecular dynamic simulations is available on a variety of platforms, including CHARMM, GROMACS, NAMD, AMBER, etc. We use AMBER (Chapter 4) for MD simulation analysis.

2.3.10. MM-PBSA analysis

The binding free energy of the complex formation was computed using the molecular mechanics Poisson-Boltzmann surface area (MM/PBSA) approach to provide additional insight into the stability of the docked complexes[22]. The binding free energy between protein-protein systems[23], protein-protein systems with inhibitors [23], protein-nucleic acid complexes [24], and protein-carbohydrate systems [25] is frequently calculated using the MM/PBSA. The following equations are used in the MM/PBSA scheme to determine the free binding energies:

$$\Delta Gbind = \Delta H - T\Delta S \approx \Delta Einternal + \Delta Gsolv - T\Delta S$$

$$\Delta Einternal = \Delta Ecovalent + \Delta Eelec + \Delta EvdW$$

$$\Delta Gsolv = \Delta Gpol + \Delta Gnon - pol$$

Where ΔEinternal, ΔGsolv, and $T\Delta S$ represent the total internal energy, solvation-free energy, and conformational entropy, respectively. For the solute and solvent, respectively, the dielectric constants were adjusted to 1.0 and 80.0, which stand for a material's propensity to contain an electric flux. It is described as the ratio of the substance's permittivity to the vacuum's permittivity. Water is known to have a high dielectric constant compared to non-polar systems and bio-macromolecules, which illustrates the effectiveness of water as a solvent. The developing aggregation in the crystals is decreased by the high dielectric constant of water. The key residues were responsible for complex formation were identified for each complex system by performing the per-residue decomposition using the MM-GBSA approach [26].

Chapter 3

Identifying Conserved G-quadruplex Motifs in *Monkeypox Virus*

3.1. Introduction

Monkeypox virus (MPXV), belongs to the Poxviridae family of genus orthopoxvirus, possesses a double-stranded DNA genome with lipoprotein to envelop and size of range between 200-250 nm, causes diseases in animals and humans when came in contact with infection (zoonotic diseases). MPXV completes its lifecycle in the cytoplasm instead of a nucleus of infected cells[1,2].

In Denmark, the first isolate of the monkeypox virus was identified in 1958 from the monkey. The first human case was identified in the Democratic Republic of Congo, in 1970 and the MPXV outbreak outside of the African continent was first reported in the U.S.,2003which takes global attention to the monkeypox virus. Currently, this virus has two clades one is Central African and another one is West African clade. Central African clade has more fertility rate compared to the West African clade [2,3].

Monkeypox comes under a self-limited disease (which can resolve spontaneously) with symptoms lasting from 2 to 4 weeks. The case fatality ratio (CFR) has been around 0.62 - 9.51% and estimated CFR 4.14% and the analysis of the CFR trend is decreasing from the previous year. The monkeypox virus infestation can be divided into two periods, the invasion period (lasts between 0-5 days) and the skin discharge period (within 1-3 days of the appearance of fever). First, the symptoms start with a high body temperature and shivering intense muscle and headache, and swelling of the lymph nodes. In the skin discharge period, all the body parts which are associated with the discharge get affected like the face in 95% of cases, feet in 75% of cases, oral mucous membrane, genitalia, cornea, and conjunctivae.

Antigen and Antibody tests are not recommended due tocross- reactivity shown by orthopoxvirus. Therefore, polymerase chain reaction (PCR)is the most recommended test for monkeypox-specific confirmation [4].

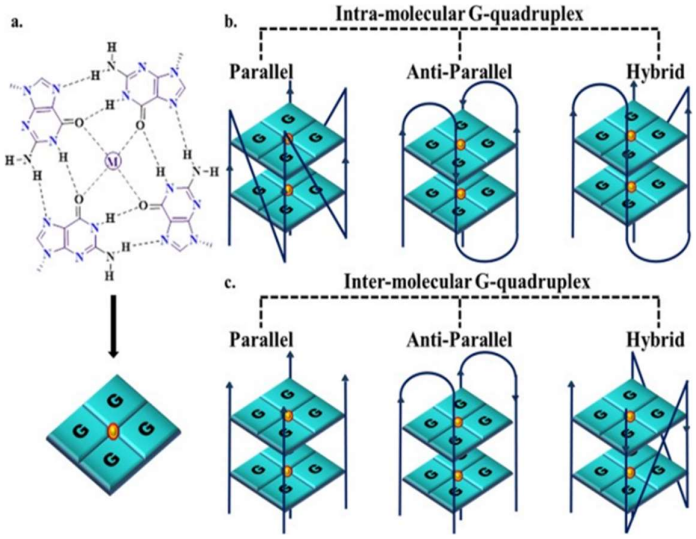
From January to November 23, 2022, a total of 80,899 laboratory-confirmed cases and 3,063 probable cases, including 26 deaths, have been reported to WHO. Currently, there is no valid vaccine or drugs specifically for monkeypox virus treatment. The smallpox vaccine has been approved by WHO for use in monkeypox treatment as different studies indicated that it is 85% effective. Therefore, specific therapeutic strategies are needed to counter even the emerging strains. As the orthopoxviridae family is serologically cross-reactive, conventional methods of using whole protein can't be effective, Small antigenic fragments (epitopes) can be used to solve the problem.

The lack of effective drugs and vaccines against the monkeypox virus has raised the importance of this topic to search for novel strategies to prevent its future outbreak. Therefore, various studies have been started targeting a nucleic acid structure that affects its protein binding affinity and looking for other alternate strategies to make anti-compound against monkeypox.

G-quadruplexes are non-B secondary structures that are formed by guanine-rich DNA or RNA segments. The Hoogsteen hydrogen bonds allow the guanine residues to form a quartet, which then stack on top of one another to form the entire G-quadruplex structure. [6,7] Cations, such as potassium or sodium ions, stabilize the structure by balancing out the negative charges on the carbonyl groups of the guanine residues. The stability and topology of G-quadruplex structures are influenced by several factors, including the size and number of the individual G-tracts, the length and composition of the loops connecting the G-tracts, and various physiological factors.

The presence of pockets in G-quadruplex structures allows small molecules to intercalate or bind to the structure, increasing its structural stability. This feature has led to the development of potential anti-viral drugs that target G-quadruplexes in the genome of certain viruses, including the human immunodeficiency virus (HIV) and the hepatitis C virus (HCV). These drugs

are designed to selectively stabilize the G-quadruplexes in the viral genome, disrupting their replication and preventing the spread of the virus.



.

Figure 3.1. (a) The G-quartet structure formed by Hoogsteen base pairing of four guanine residues. (b) The intra-molecular topology of the G-quadruplex structure forms parallel, anti-parallel, and hybrid conformations. (c) The inter-molecular topology of the G-quadruplex structure forms parallel, anti-parallel, and hybrid conformations. Adapted from the (DOI: 10.1038/s41598-020-58406-8)

Several experimental antivirals have been shown to be effective against MPXV These include the antiviral cidofovir and its analog brincidofovir, which work by obstructing viral DNA polymerase, and tecovirimat, which blocks the development of the virus extracellular envelope and thus prevents cell-to-cell transmission. The MAV vaccine and modified attenuated VACV (Ankara strain) were both approved by the USFDA and EU in 2019 for the protection of monkeypox, but none of them underwent clinical trials for evaluation.

In guanine-rich regions, the secondary nucleic acid structure known as the G-quadruplex (GQ) is prevalent. Numerous human viruses have large amounts of potential quadruplex-forming sequences (PQS) encoded in their genomes. For example, Braco-19 inhibits the development of EBV nuclear antigen-1 and suppresses the hepatitis C virus. PQS could be targeted for viral suppression and biomedical diagnostics. [27-30].

To identify the sequences that could form G-quadruplexes in the monkeypox genome. We used the QGRS mapper and G4IPDB to analyze the MPXV genome. We use a variety of techniques to further characterize the sequences, including NMR spectroscopy, CD spectroscopy, fluorescent binding analysis, UV melting, native EMSA, and Isothermal calorimetry.

3.2. Materials and methods

3.2.1. Prediction, conserved motif enrichment, and functional analysis of G-quadruplex motifs in *Monkeypox virus* strains

MPXV strains that have been fully sequenced have been downloaded from the National Centre for Biotechnology Information (NCBI). Using our previously created G-quadruplex predictor program, these strains were then thoroughly mined for the probable G-quadruplex motifs in both the sense and antisense strands [34]. The aforementioned regular expression was employed by this prediction programme.

G{T1}[X]{L1}G{T1}[X]{L2}G{T1}[X]{L3}G{T1}

Where X is any nucleotide (A, T, G, or C), L1, L2, and L3 represent the variable loop region and can be any number from 1 to 7, T1 represents consecutive tracts of guanine and can be any number from 2 to 7. We employed G-tracts of 3 or 4 and loop lengths of 1–20 nucleotides for our prediction[34N].

3.2.2. Oligonucleotide preparation for CD and ITC analysis

The predicted G4 oligonucleotide sequences were supplied by Integrated DNA Technologies (Iowa, United States) (Table 3.1). The manufacturer's instructions were followed to prepare 100 M stock solutions. Oligonucleotides were re-annealed by heating at 95 ºC for 10 minutes and slowly cooling at room temperature for 2 hours before each series of experiments. All of these oligonucleotides were dissolved in four independent 50 mM solutions of K+, Na+, Li+, and Mg2+ in four different Tris-buffers (pH=7.0, 10mM).

3.2.3. CD Spectroscopy and melting analysis

With a Jasco J-185 Spectropolarimeter (Jasco Hachioji, Tokyo, Japan) and a scanning speed of 20 nm/min from 220 to 320 nm, CD experiments were carried out for each oligonucleotide in various buffer conditions at 25 ºC. There was a Peltier Junction temperature (PTC-423S/15) controller in the instrument. In four consecutive studies, spectra were taken in a cuvette with a 1 mm path length and a final concentration of 10 µM of all oligonucleotides in Tris buffer (10 mM, pH 7.4) with 50 mM concentrations of K+, Na+, Mg2+, and Li+ and with a sample volumeof200µl. Before each spectrum recording, a blank spectrum was taken to prevent signal contribution from the buffer.

Each MPXV GQ sequence was subjected to melting analysis (final concentration of 20 µM) at temperatures ranging from 25 ºC to 98 ºC in buffers containing 50 mM of each of four distinct cations: K+, Na+, Mg2+, and Li+. For each of the melting studies, the heating rate was set at 1 °C/min.

Table 3.1. List of MPXV-GQs along with the other sequences used in this chapter

S.No	Name	Template Sequence (5' – 3')
1.	B3R	TAGGAAAAGGTGGATTCGGTA
2.	DUTPase	GAGGAGATCAAGGGTTTGGATCAACAGGAC
3.	B12R	AAGGAGGAGAACATGGATAAGGTT
4.	A25R	ACGGATGGGATTCTTCGGTAACTTGGAG
5.	I4L	ATGGCTATTCTGGCGGCTAGAATGGCA
6.	L6R	ATGGTAGACTAGGTGCTATGGATGGGGCA
7.	F4L	TAGGTGATGCTGGTTCTGGAGATTCTGGAG
8.	EEVmaturationProtein	ACGGTGCTAGGGTTATTTGGATGGAT
9.	ESR	ATGGTGTATTGGACCTGGTAGACGGAA
10.	IMVMembrane Protein	CAGGTTGTTGGGGTTGGTATTGGTT

11.	A19R	CAGGTAGAGGCAGTCGGATTGGAA
12.	A49R	ACGGTTACTGGACCAGTGGGGCAACTGTGGAT
13.	c-myc	TGAGGGTGGTGAGGGTGGGGAAGG

3.2.4. Electrophoretic Mobility Shift and Gel Retardation Assay

With 20% polyacrylamide gel and 1X TBE buffer, native PAGE was performed. Each sample was dissolved in a separate solution of Tris buffer (pH=7, 10 mM, four distinct cations: K+, Na+, Mg2+, and Li+, each at 50 mM). An oligonucleotide of comparable length (G nucleotide replaced with T) was used as a negative control for each GQ sequence, and standard G-quadruplex (c-myc and Bcl-2 DNA) was used as a positive reference. Each oligonucleotide sample was loaded onto a vertical gel unit system with 20 µl, and electrophoresis was carried out at 40 ºC and 90 Volts. The gels were examined on the ImageQuant LAS 4000 gel doc (GE Healthcare Biosciences Ltd., Sweden) after being stained with ethidium bromide to visualize them.

For the gel retardation experiment. For 30 min at 25 °C, 30 µM MPXV G4 DNA was incubated with different Braco-19 concentrations from 0.00 µM to 100 µM. The incubation results were then resolved on a 20% native polyacrylamide gel [29:1 acrylamide/bicarbonate]. The EtBr (Ethidium bromide) solution was used as a staining solution for the native page Gel and images were analyzed with the help of Image Quant LAS (GE health care, Bioscience Ltd, Sweden).

3.2.5. One dimensional ^{1}H-NMR Spectroscopy

NMR spectroscopy research was carried out using a 5 mm broadband inverse probe by AVANCE 500 MHz BioSpin International AG in Switzerland. In a buffer solution of potassium phosphate 50 mM, the MPXV-GQ sequences were dissolved. H2O/D2O was used as the solvent in all NMR experiments at a 9:1 ratio. We employed *sodium trimethylsilylpropanesulfonate* salt (DSS) as an internal reference at a temperature of 298 K with a spectrum width of 20 ppm. Topspin 1.3 software was used for the interpreting, integration, and evaluation of NMR data.

For analyzing the interaction of Braco-19 with the MPXV-GQs, an NMR broadening analysis was performed. The MPXV-GQ oligonucleotides were gradually added into the Braco-19 (200 μM) solution, and the corresponding spectra were recorded.

3.2.6. Fluorescent Binding Assay

The fluorescence titration experiment was performed using the Synergy H1 multi-mode plate reader. 5 nM of ligand was added to each of the 12 wells of a 96-well plate and gradually the target DNA was titrated in each of the wells (0–20 mM). Precisely, the highest concentration of DNA, i.e., 20 mM, was added to the first well, followed by dilution in the subsequent well; for example, the second well then contained 10 mM, the third well contained 5 mM, and so on until the eleventh well, while no DNA was added in the last well (twelfth well) serving as blank. The enhancement in emission fluorescence was observed at specific excitation of the ligand molecule (λex = 360 nm and λem = 440 nm for Braco-19) on the addition of DNA and the change in emission fluorescence (DF) for all datasets was normalized to the last well (containing no DNA). The two-site saturation using global curve fitting was used to extrapolate the curve between DF and concentration of DNA (mM) using SigmaPlot 12.5 software (Systat Software, Chicago, IL, USA) according to the following equation that accounts for two receptor binding sites with two different affinities, K_D^1 and K_D^2:

$$y = Bmax1 * abs(x) / K_D^1 * abs(x) + Bmax1 * abs(x) / K_D^2 * abs(x)$$

where Bmax is the maximum number of binding sites and KD is the dissociation equilibrium constant. Each binding study was performed in duplicate at room temperature (RT) and duplex DNA was used as the control.

2.3.7 UV-Visible spectrophotometer melting analysis

The melting analysis of oligonucleotide sequences containing G-quadruplexes is done to check the thermal stability as GQs in the sequence help to stabilize the structure in the presence of cations that insert in the GQ pocket. With a PerkinElmer LAMBDA™ 25/35/45 UV/Vis Spectrophotometers, UV melting experiments were carried out for each nucleotide in increasing K+ concentration to check thermal stability and there was a Peltier Junction temperature (PTC-

423S/15) controller in the instrument. In the experiment, four consecutive analysis was done by increasing K+ concentration from 0 Mm to 200 mM, spectra were taken in a cuvette with a 1 mm path length. A final concentration of 30 μM of an oligonucleotide is added in 10x potassium phosphate buffer with a sample volume of 300 μL. Before each experiment, a blank spectrum was taken to prevent signal interference from the buffer.

3.2.8. Isothermal titration calorimetric (ITC) analysis

Using a MicroCal iTC200 calorimeter (GE Healthcare, Biosciences Ltd., Sweden), ITC analysis was done on all three PGQs as well as a linear DNA (negative control). The total amount of each of the oligonucleotides was 20 μM and they were all dissolved in a buffer of potassium phosphate 50 mM. The oligonucleotide was applied in the ITC cell, and 200 M of BRACO-19 dissolved in the same quantity of buffer were titrated 21 times by using a syringe, with the initial injection of 0.4 L followed by 1.8 L of the ligand at each step for the period of 3.6 s and the decay of 90 s between each step. The same quantity of BRACO-19 was added to the 50 mM potassium buffer to calculate the heat of dilution of the oligonucleotides, which was then deducted from the interaction isotherms of the oligonucleotides preceding the curve fitting analysis. Data were collected in three separate instances and analyzed using Origin Scientific Software version 7 (Microcal Software Inc., Northampton, MA, USA) to determine the association constant by its two-site binding mode analysis.

3.3. Result and Discussion

3.3.1. Monkeypox Virus genome contains conserved G-quadruplexes

The tenacious surge for developing novel anti-microbial approaches to combat multi-drug resistance has expanded the use of evolutionary conserved nucleic acids as drug targets. The G-quadruplex motifs in human genomes are one of the vast studied non-canonical structures, but their presence in human pathogens is quite captivating and yet unexplored. The varied genomic content in viruses and their potential to cause lethal diseases in humans have drained the scientific community's attention towards the presence of G-quadruplex structures in these pathogens.

G-quadruplex structures are created by the nucleotide sequence that consists of two or more successive guanine residues spaced by an intermediate region. However, only two G-tetrads are required to form the G-quadruplex structures, as shown in viruses like the Florida manatee species infecting papillomavirus and the pseudorabies virus [31,32]. Using our earlier published G-quadruplex prediction tool, we conducted the genome-wide screening for prospective G-quadruplex forming sequences. Other online tools like QGRS mapper and QuadBase 2 reiterated the predictions further, and we listed 12 GQ-forming sequences throughout the MPXV genome, the analysis utilized bioinformatics where the optimum length being up to 30 nucleotides and the smallest number of guanine repeats was kept at two, with a loop size of 1 to 7 nucleotides. As a result, it was projected that these GQs would conserve often among the various viral isolates, and this information is listed in *(Table 3.1)*.

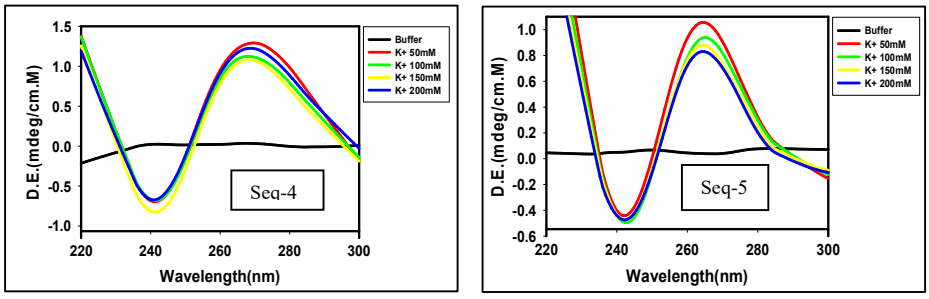
Table 3.2. List of GQs with G-tract ≥2 and loop length 1≤ L ≥ 7 and conserved in 98% Monkeypox Virus along with their gene name.

S.No	Name	G4 motifs	Conservation	Gene
01	MPV_Seq2	TAGGAAAAGGTGGATTCGGTA	> 99%	B3R
02	MPV_Seq4	GAGGAGATCAAGGGTTTGGATCAACAGGAC	> 99%	dUTPase
03	MPV_Seq5	AAGGAGGAGAACATGGATAAGGTT	> 99%	B12R
04	MPV_Seq8	ACGGATGGGATTCTTCGGTAACTTGGAG	> 99%	A25R
05	MPV_Seq10	ATGGCTATTCTGGCGGCTAGAATGGCA	> 99%	I4L
06	MPV_Seq11	ATGGTAGACTAGGTGCTATGGATGGGGCA	> 99%	L6R
07	MPV_Seq12	TAGGTGATGCTGGTTCTGGAGATTCTGGAG	> 99%	F4L
08	MPV_Seq13	ACGGTGCTAGGGTTATTTGGATGGAT	> 99%	EEV maturation Protein
09	MPV_Seq15	ATGGTGTATTGGACCTGGTAGACGGAA	> 99%	E5R
10	MPV_Seq16	CAGGTTGTTGGGGTTGGTATTGGTT	> 99%	IMV Membrane Protein
11	MPV_Seq21	CAGGTAGAGGCAGTCGGATTGGAA	> 98%	A19R
12	MPV_Seq22	ACGGTTACTGGACCAGTGGGGCAACTGTGGAT	>98%	A49R

.

3.3.2. Evaluating the topology and stability of the GQs using Circular Dichroism

One of the popular methods for examining the topology of the G-quadruplex structure is circular dichroism. According to the sequence, length of the loop, and bound cation, the G-quadruplex may acquire a parallel, anti-parallel, or hybrid conformation. Parallel G-quadruplex topology is indicated by a positive peak at ~260 nm and a negative peak at ~240 nm. A positive peak at ~290 nm and a negative peak at ~260 nm, however, indicate an anti-parallel G-quadruplex structure. The hybrid structure is shown by two positive peaks at ~260 nm and ~290 nm and a negative peak at 240 nm. Furthermore, a positive peak is formed by all G-quadruplex conformers at 210 nm [33]. Potassium ion is known to stabilize the G-quadruplex structures better because of their atomic size and ionic radii [34]. The increasing ellipticity with increasing K+ concentration depicts the higher stability of these secondary structures. In sequences 13 and 16 with the gradual increase in K+ concentration, from 50mM to 200mM, the ellipticity of the CD spectra was observed to increase markedly in both the GQs but with sequences 5, 10, 11, 12, 15, 21, 22 ellipticity of the CD spectra was observed to decrease as the k+ concentration was increased *(Figure3.2.)*



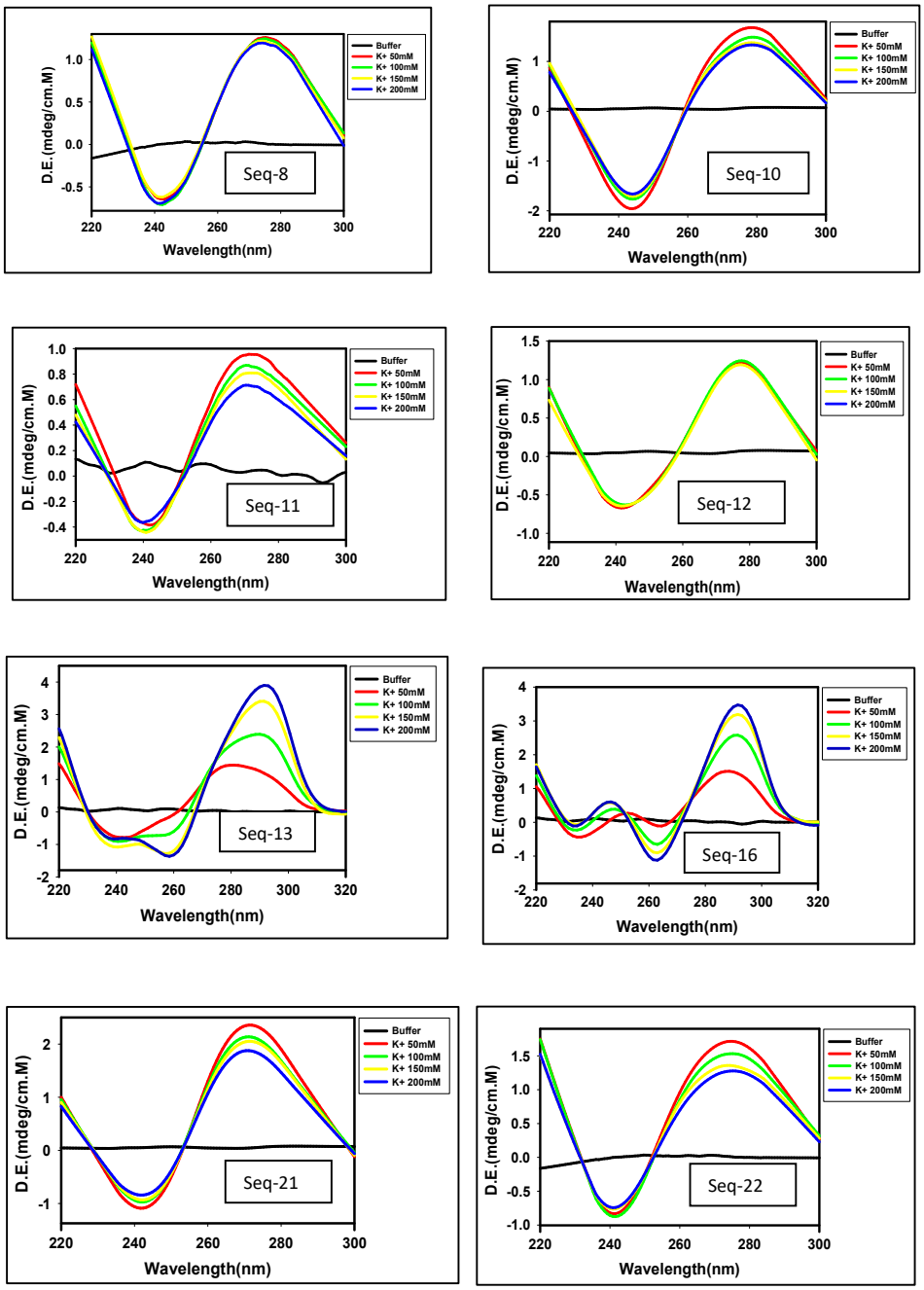


Figure 3.2. Shows CD spectra analysis of all sequences with increasing K+ ion concentration in the presence of Tris-Cl buffer (10 mM).

The durability of the G-quadruplex structure is differently impacted by various cations. Following is a ranking of some well-studied cations' stabilizing power: Li+ follows K+, Na+, Mg2+, and Mg. Therefore, in these four different cation-containing buffers (K+, Na+, Li+, and Mg2+), we carried out the CD spectroscopy of MPXV-GQs. Sequences 13 and 16 were found to exhibit anti-parallel topology*(Figure 3.2 and Figure 3.3).* Whereas all other sequences showed to bear parallel topology and we explored the stability of the putative G- quadruplex with an increase in the K+ion concentration from 50 mM to 100 mM. CD melting studies of sequence 13 and sequence 16 also confirmed that the GQ structures are stabilized following the addition of Braco-19 (higher Tm value in D/N = 1) as compared to the GQs alone. The stabilization of the G-quadruplex structures makes it difficult to unfold, thus leading to an increase in melting temperature. A higher Tm was also observed when the interactions of these GQs with Braco-19 were evaluated with the use of CD melting experiments *(Figure 3.3 and Figure 3.4).*

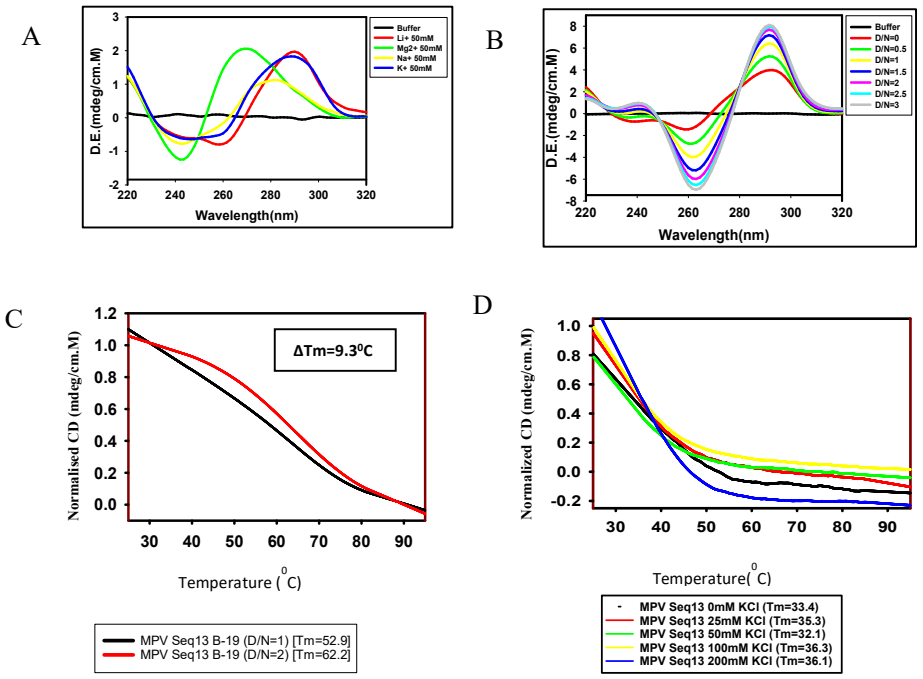


Figure 3.3. CD spectra analysis of MPXV sequence 13. (A) With different cations (B) with Braco-19 in different D/N ratio and CD melting analysis (C) with Braco-19 (D)with increasing K+ concentration.

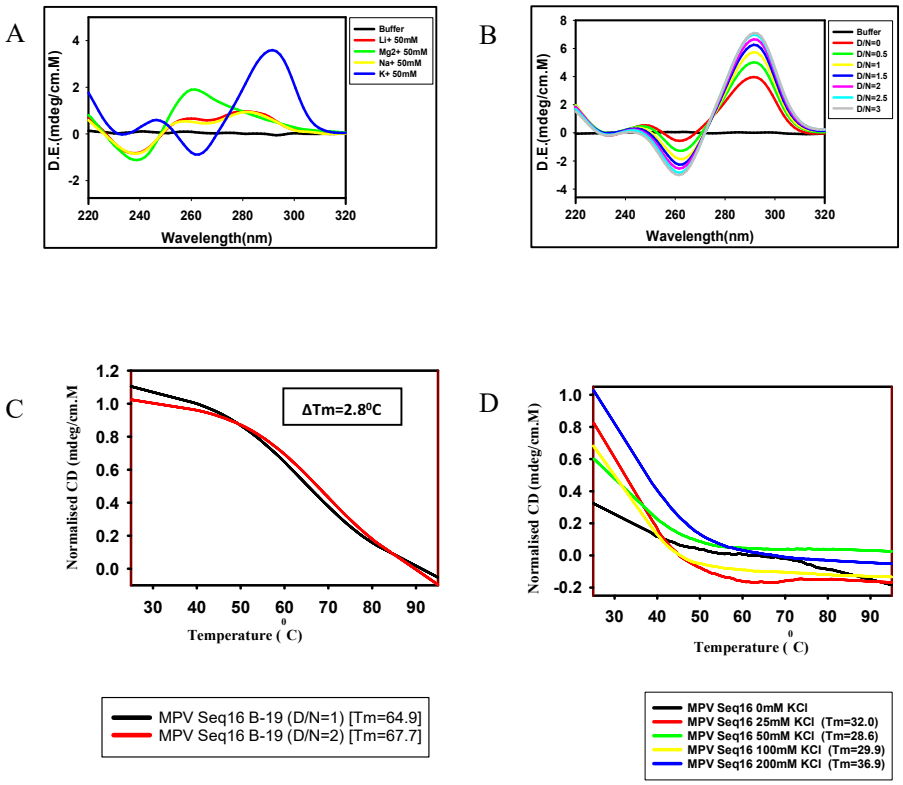


Figure 3.4. CD spectra analysis of MPXV sequence 16. (A) With different cations (B) with Braco-19 in different D/N ratios and CD melting analysis (C) with Braco-19 (D)with increasing K+ concentration.

3.3.3 Nuclear Magnetic Resonance (NMR) confirmed the formation of G-quadruplex structures in the MPXV motifs.

The most trustworthy method for verifying the development of the G-quadruplex structure by the nucleic acid sequences is NMR spectroscopy. In order to confirm that MPXV-GQs do really form the G-quadruplex structure, 1D 1H NMR spectroscopy was used. Hoogsteen base pairing can be seen in the typical G-tetrads of the G-quadruplex structure as a chemical shift in the range of 10–12 ppm in 1D 1H NMR spectra, whereas canonical G–C Watson–Crick base pairing can be seen as a chemical shift in the range of 12–14 ppm. The creation of the G-quadruplex structure was confirmed by three MPXV-GQs (13, 16, 22), which had an imino proton resonance between 10 and 12 ppm (Figure 3.5). and sequence 10 spectra analysis, the peak is present between a range of 12-14 ppm which shows the Watson-Crick pairing i.e., maybe a hairpin loop or any other secondary structure will be present. All the other sequences do not show any characteristics peak.

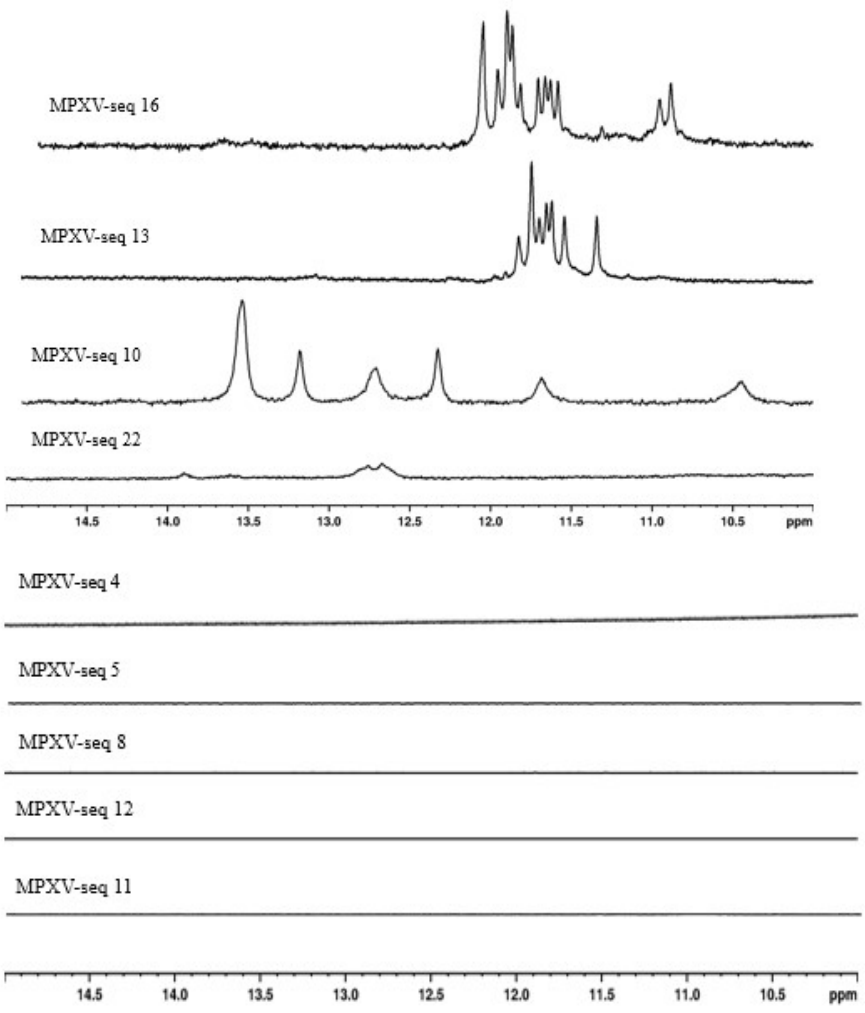


Figure 3.5. NMR Spectra. 1D ^{1}H NMR spectra of all the MPXV sequences in the presence of K+ Buffer.

NMR chemical shift disruptions and peak broadening in the 1D NMR spectra of the ligand on the addition of GQ DNA can provide an insight into the mode of interaction and ligand's proton involvement in the complex formation. Considering the above facts, a series of Braco-19 NMR spectra were taken at varying concentrations of GQ DNA, and the ligand 1H-peaks were analyzed Braco-19 N,N'-(9-{[4-(dimethylamino)phenyl]amino}acridine-3,6-diyl)bis(3-

pyrrolidin-1-ylpropanamide) have various aromatic and aliphatic protons that can participate in ligand-GQ formation. A total of three types of aromatic hydrogens were present in Braco-19 (H1, and H3), while tetratosylate salt has two different aromatic protons (H2, H4, & H5) (*Figure 4.26a*). On titrating the MPXV-GQs sequence 13 and 16 in the Braco-19 solution, peaks of aromatic protons show perturbations, (*Figure 3.6- 3.7*). The involvement of aromatic hydrogens showed the π-π interaction between the GQs and Braco-19, leading to stable complex formation.

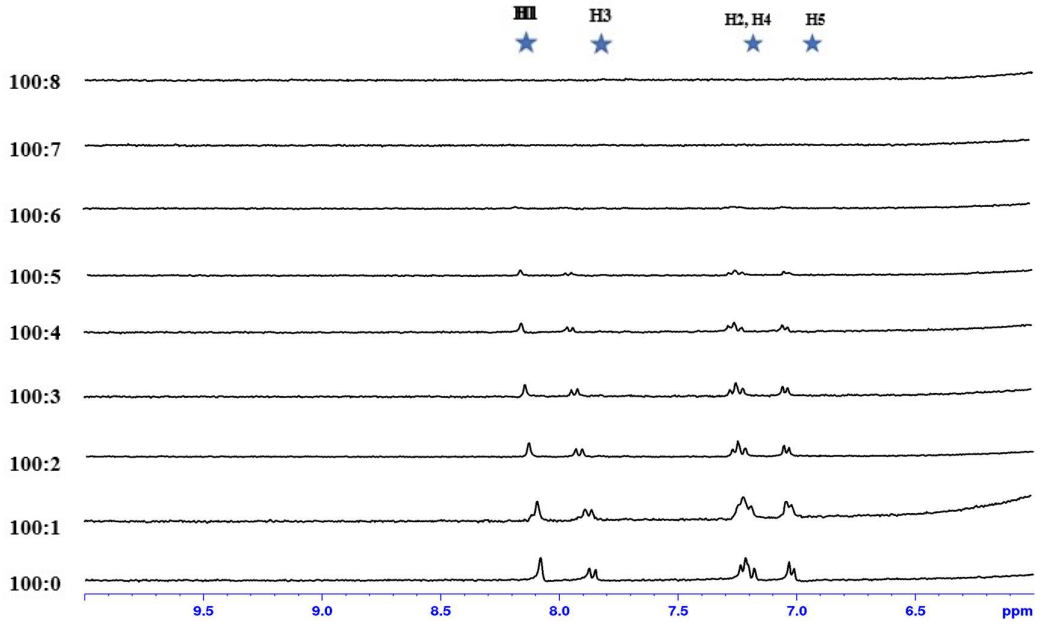


Figure 3.6. NMR peak broadening further affirms the interaction of Braco-19 with the MPXV-GQ sequence 13.

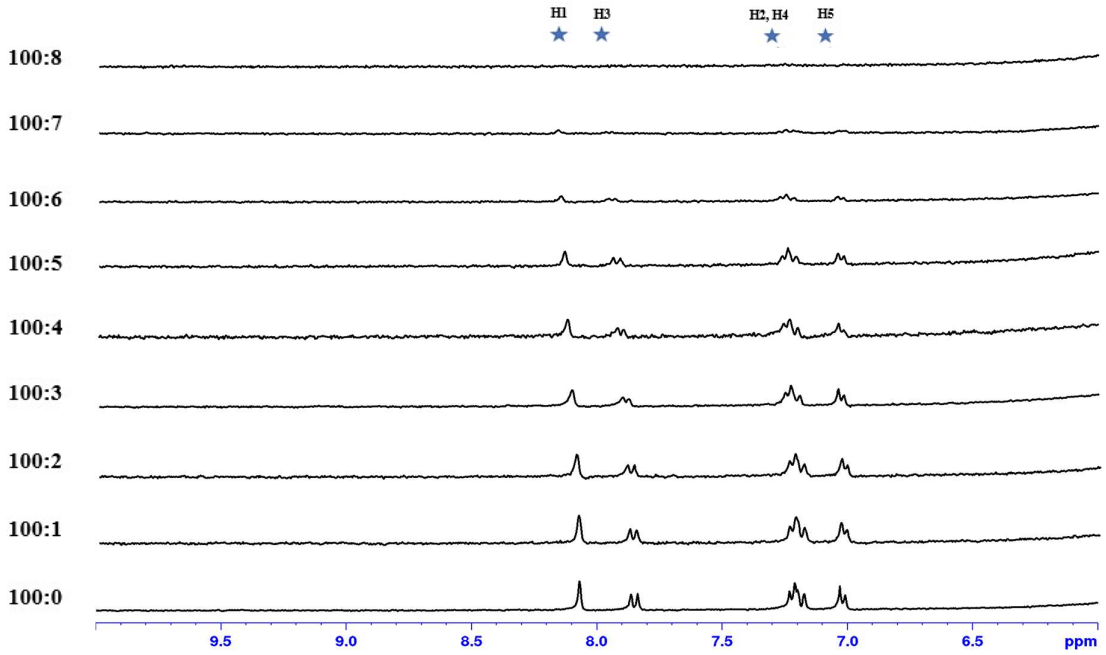


Figure 3.7. NMR peak broadening further affirms the interaction of Braco-19 with the MPXV-GQ sequence 16.

3.3.4. Electrophoretic Mobility Shift Assay (EMSA) and Gel retardation assay analysis

The gel retardation assay is a fast and efficient way to analyze the interaction of small molecules with DNA. Therefore, to study the binding of Braco-19 with MPXV DNA sequences, we conducted a gel retardation assay by incubating the MPXV DNA with different concentrations of Braco-19 (0.00–100.00 μM) for half an hour and then resolved on 20% native PAGE. We have observed the significant retardation in the band mobility by increasing the concentration of Braco-19in some of the sequences which shows strong MPXV and Braco-19 analogs complex. Braco-19 Shows significant and maximum retardation with almost all the sequences and less interaction with sequences (12, 21). Furthermore, we have also performed the gel

retardation assay of Braco-19 with mutant MPXV G4 sequence that is unable to form a G4 structure, but no significant band shift was observed (Figure 3.9).

EMSA is widely used as an experimental tool to assess the molecularity (inter or intra-molecular) of G-quadruplex structures. The intra-molecular G-quadruplex structures form a compact topology and run faster in the gel as compared to their linear counterparts. In the case of inter-molecular G-quadruplex structure, an association of more than one nucleic acid strand in G-quadruplex formation results in slower mobility. EMSA of DNA sequences 13 and 16 shows inter-molecular GQs (Figure 3.8).

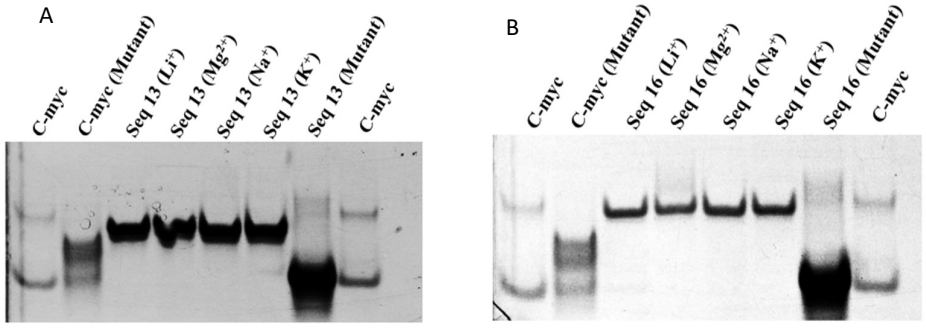


Figure 3.8. Electrophoretic Mobility Shift assay

Gel images of MPXV-GQ sequence A) 13 and B)16.

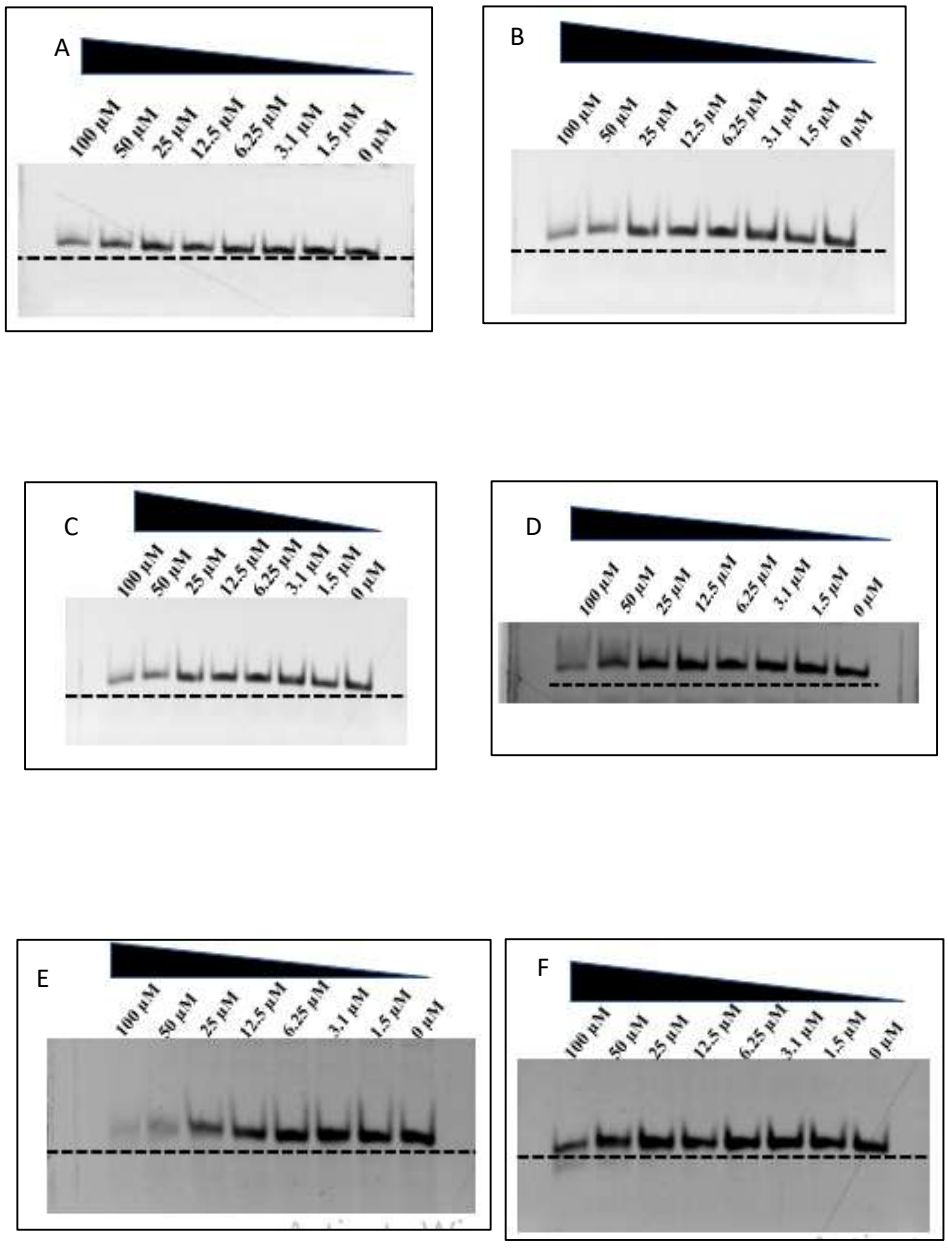


Figure 3.9. Gel Retardation assay.

Gel images of (A) MPXV-seq 4, (B) MPXV-seq 5, (C) MPXV-seq 8, (D) MPXV-seq 10, (E) MPXV-seq 11, and (F) MPXV-seq 12.

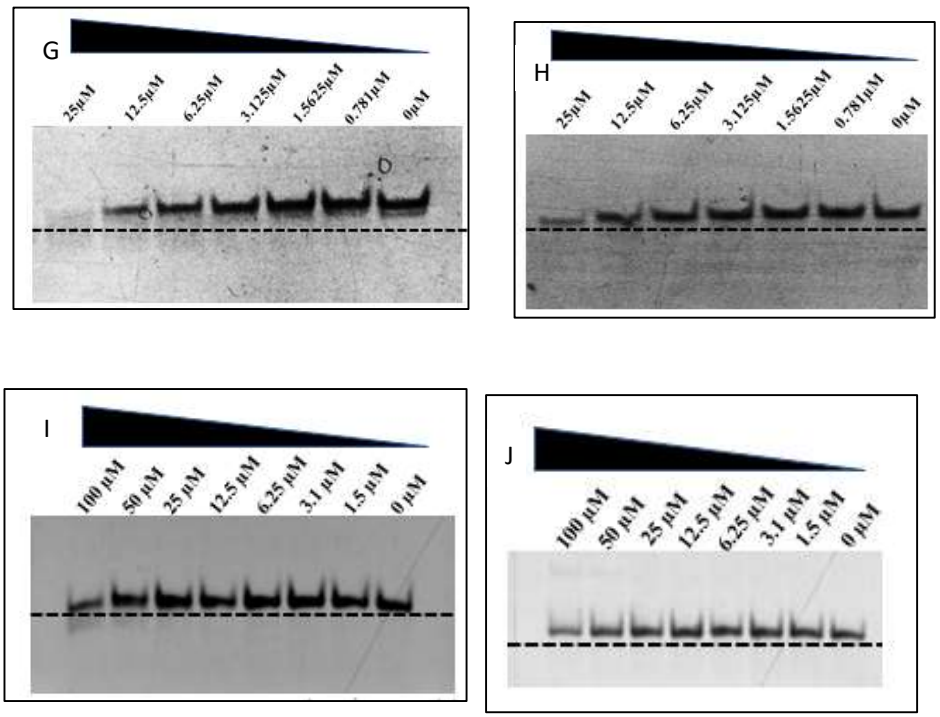


Figure 3.9. Gel Retardation assay.

Gel images of (G) MPXV-seq 13, (H) MPXV-seq 16, (I) MPXV-seq 21, and (J) MPXV-seq 22

3.3.5 G-quadruplex ligand stabilizes the MPXV-GQs by an energetically favorable interaction.

Certain small molecules like TMPyP4, Braco-19, PDS, etc. possess a characteristic planar aromatic surface. The cationic moieties in them further enhance their interaction with the G-quadruplex structures. Isothermal titration calorimetry (ITC) is one of the most trustworthy methods for analyzing molecular interactions since it depends on how enthalpy (H) and entropy (S) change when two molecules contact. Here, we conducted the ITC experiment in 50mM potassium phosphate buffer and fitted the resulting thermogram in the two-site binding model to test the specificity and affinity of BRACO-19 and with the MPXV-GQs. We used duplex DNA as a control. The

thermodynamically advantageous interaction between BRACO-19 and MPXV-GQs and the resulting development of a stable complex was shown by a negative shift in Gibbs free energy (G) (*Figures 3.7*).

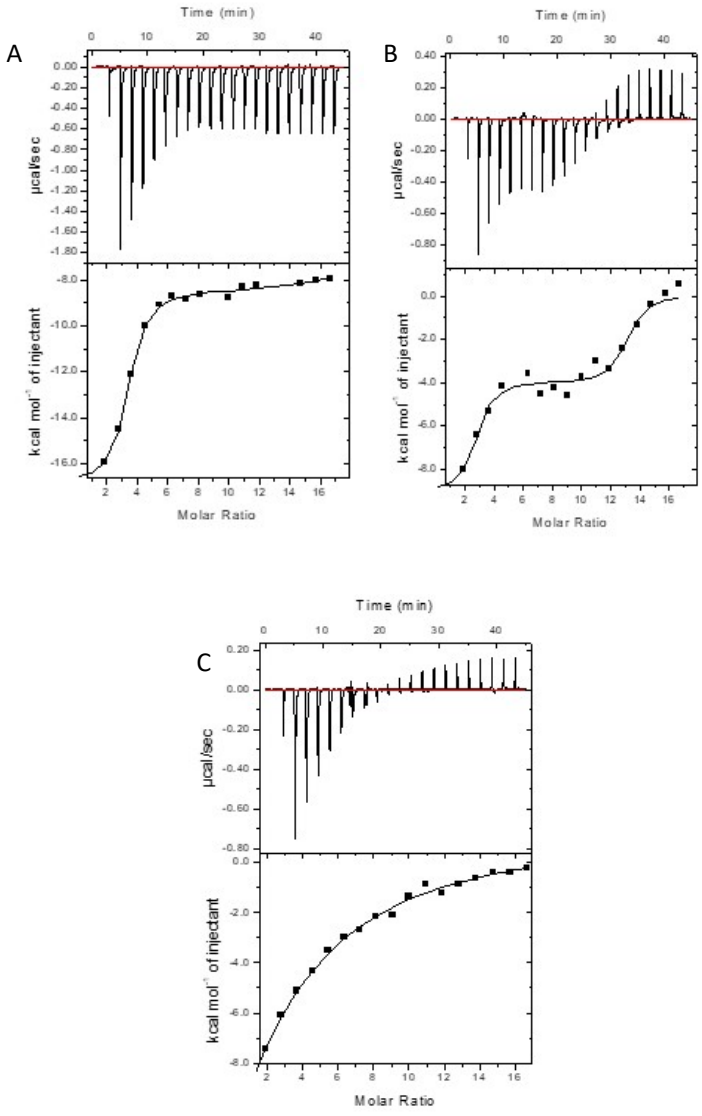
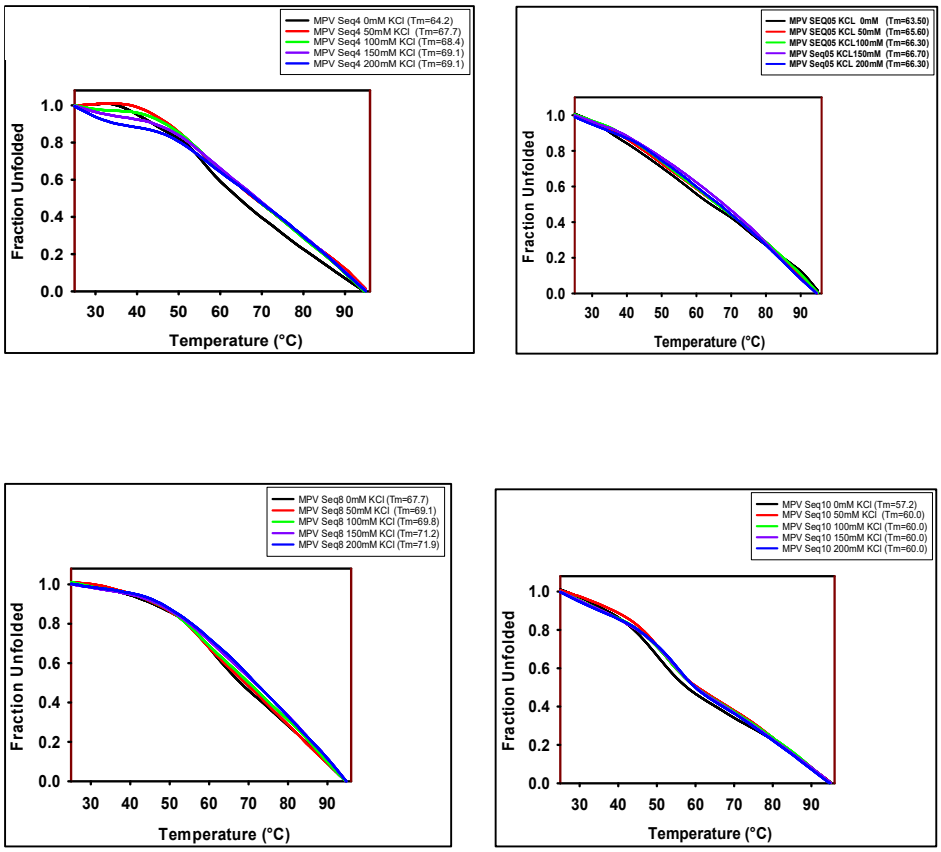


Figure 3.10. Isothermal curves of MPXV-GQ with Braco-19.

Isotherms obtained by isothermal titration calorimetric analysis of the conserved MPXV-GQs with Braco-19 as a ligand. Duplex DNA was used as a negative control. ITC curves shown in (A) MPXV-GQ 13 (B) MPXV-GQ 16 (C) Negative control.

3.3.6. Cations stabilize the MPXV-GQs by increasing thermal stability.

The melting analysis of oligonucleotide sequences containing G-quadruplexes is done to check the thermal stability as GQs in the sequence help to stabilize the structure in the presence of cations which insert in the GQ pocket. With the increases in K+ concentration, the stability of MPXV sequences also increases. All the sequences show thermal stability with an increase in potassium concentration as depicted in *(Figure 3.8).*



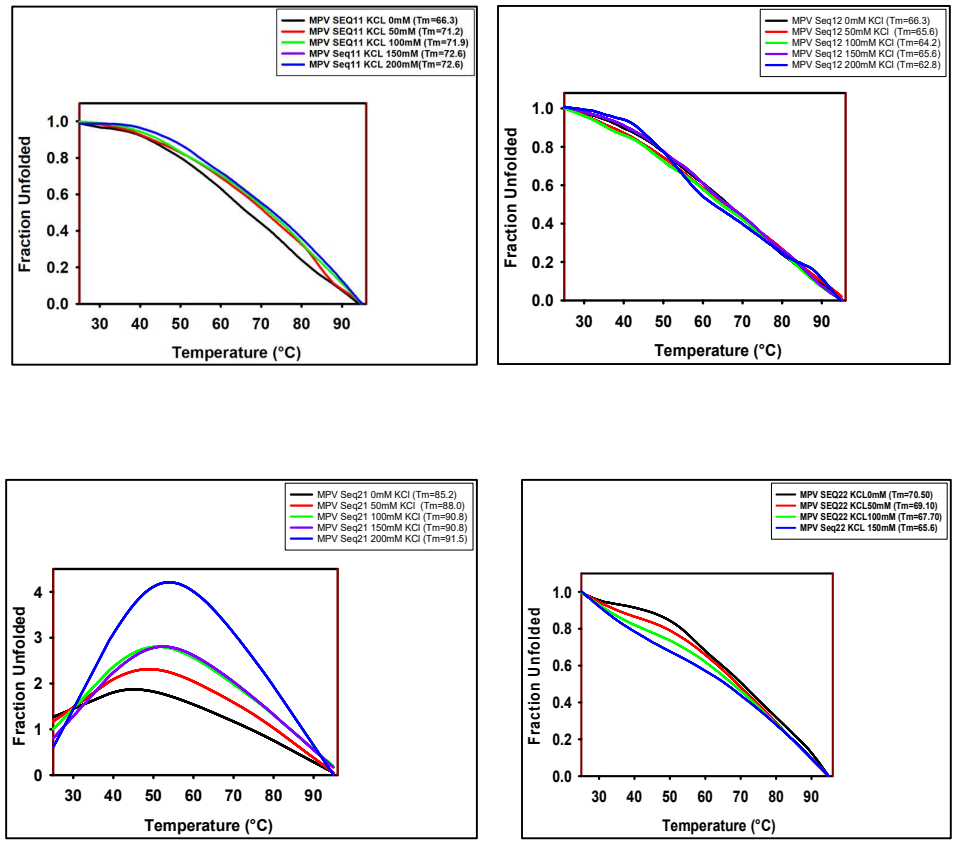
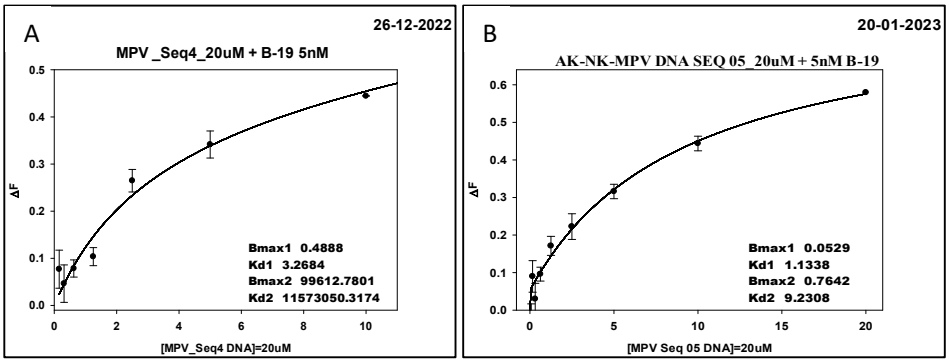


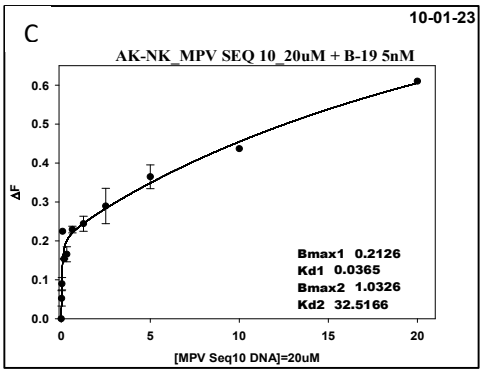
Figure 3.11. UV Melting analysis of MPXV-GQ with increasing K+ concentration.

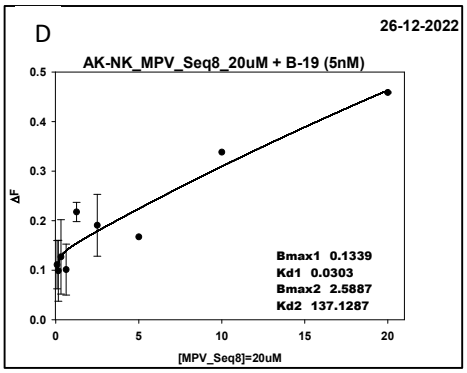
3.3.7. Interaction of MPXV-GQs with GQ-binding ligand Braco-19 demonstrated by Fluorescent Binding Assay

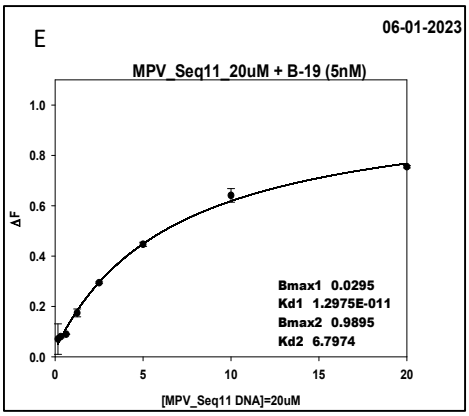
This assay is a simple optical assay that makes use of the intrinsic fluorescence property of the ligand molecule. Fluorescence-binding assay helps to determine the binding affinity between two molecules (in our case, the GQ-specific ligand, Braco-19, and the DNA) and obtain a quantitative value in the form of the dissociation constant (K_D). The binding efficiency between the ligand and the biological molecules, i.e., DNA, is evaluated through the difference between the fluorescence intensity observed before the addition of the DNA and after its interaction with the ligand. The successful interaction results in either increase (enhancement) or a decrease in fluorescence intensity (quenching). This change

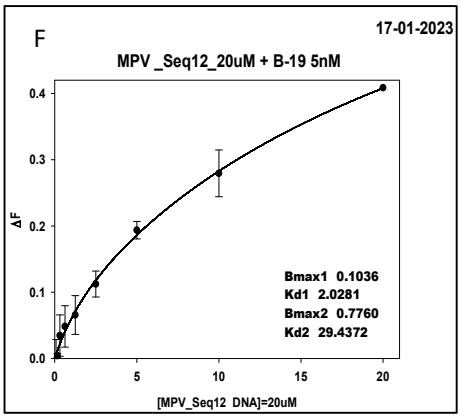
in fluorescence (ΔF) is plotted with respect to the concentration of DNA (GQs or canonical) using global curve fitting methods to obtain the K_D values for each interaction. The respective K_D value is used to analyze the binding affinity of the fluorescent molecule with each nucleic acid. In the present study, all the ZGQs showed higher binding affinity (>100-fold higher) for Braco-19 in comparison to the control DNA sequence (Figure 3.12.). The variation in binding affinity of different ZGQs with Braco-19 is probably due to different polymorphic structures formed by the GQ-sequences and so the ligand molecules bind with the different DNA sequences with varying intensity. It should be noted that a higher affinity of Braco-19 toward the GQ sequences is attributed to the π-π interactions between the aromatic region of the ligand and the GQ-forming scaffolds of the ZGQs. Moreover, Braco-19 contains two pyrrolidine rings, which contribute to the cationic nature of the molecule and help to interact with the anionic core of the GQ structures. All these properties contribute to the mode of binding between the ligand and the biomolecule and thereby can exhibit different binding modes, for example, groove binding or end-stacking modes. Thus, we evaluated the interaction ability of these GQ-specific ligands with the predicted ZGQs. In our study, all of the sequences were fitted in the two-mode binding model depicting more than one site being present in the DNA sequences for the ligand binding. The KD1 values were higher in the fitted case, confirming preferential binding mode and better affinity of the ligand toward the ZGQ sequences.

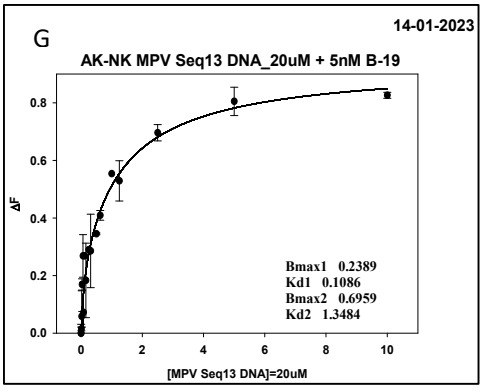


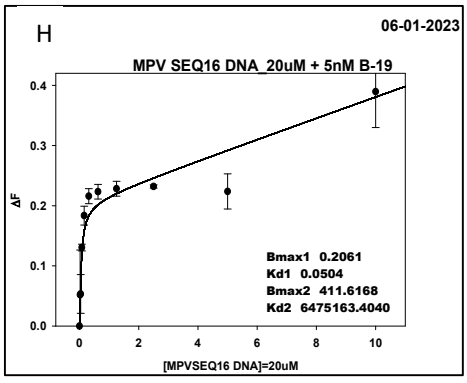

C
10-01-23
AK-NK_MPV SEQ 10_20uM + B-19 5nM
ΔF
Bmax1 0.2126
Kd1 0.0365
Bmax2 1.0326
Kd2 32.5166
[MPV Seq10 DNA]=20uM


D
26-12-2022
AK-NK_MPV_Seq8_20uM + B-19 (5nM)
ΔF
Bmax1 0.1339
Kd1 0.0303
Bmax2 2.5887
Kd2 137.1287
[MPV_Seq8]=20uM


E
06-01-2023
MPV_Seq11_20uM + B-19 (5nM)
ΔF
Bmax1 0.0295
Kd1 1.2975E-011
Bmax2 0.9895
Kd2 6.7974
[MPV_Seq11 DNA]=20uM


F
17-01-2023
MPV _Seq12_20uM + B-19 5nM
ΔF
Bmax1 0.1036
Kd1 2.0281
Bmax2 0.7760
Kd2 29.4372
[MPV_Seq12 DNA]=20uM


G
14-01-2023
AK-NK MPV Seq13 DNA_20uM + 5nM B-19
ΔF
Bmax1 0.2389
Kd1 0.1086
Bmax2 0.6959
Kd2 1.3484
[MPV Seq13 DNA]=20uM


H
06-01-2023
MPV SEQ16 DNA_20uM + 5nM B-19
ΔF
Bmax1 0.2061
Kd1 0.0504
Bmax2 411.6168
Kd2 6475163.4040
[MPVSEQ16 DNA]=20uM

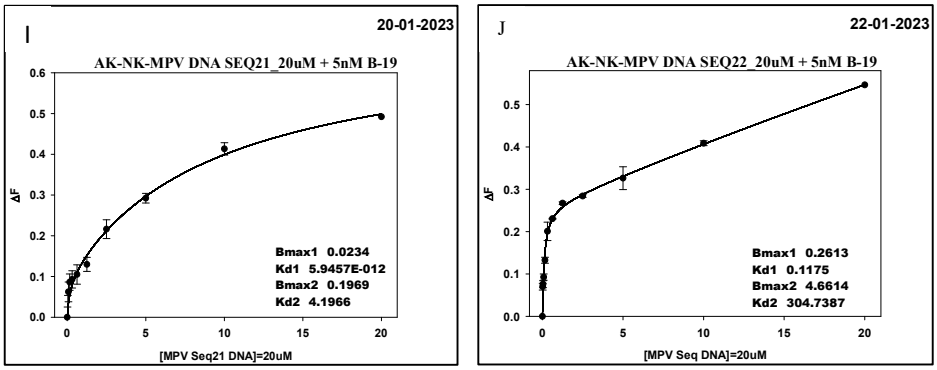


Figure 3.12. Fluorescent binding assay of different MPV sequences (A-J).

3.4. Conclusion

As a concluding remark, the present study demonstrates that crucial *Monkeypox virus* genes include stable and highly conserved G-quadruplex structures. The current study is a proof-of-concept analysis of the discovery and description of G-quadruplex motifs in *Monkeypox virus's* important genes. These MPXV-GQs sequences 13, 16, and 22 can offer a cutting-edge platform for the creation of therapeutics to combat *Monkeypox virus* infection. A planer hetero-aromatic chromophore is a structural component found in the majority of widely used G-quadruplex binders that aids in their ability to π-π interaction with the G4 motif. Acridine derivatives have been extensively used to create a variety of di- and tri-substituted, which exhibit much higher specificity for G-quadruplexes than for duplex DNA [49,50,52,59-65]. Herein, Acridine derivatives, BRACO-19 was observed to bind and provide stability to these G-quadruplex structures possessing genes, which may reduce the expression of these genes and thereby proposed as a novel G4 mediated therapeutic approach for combating the infection of *Monkeypox virus* in humans.

CHAPTER -4

An *In-silico* study for multi-epitope subunit vaccine development against *Monkeypox virus*

4.1. Introduction:

The recent outbreak of monkeypox worldwide has taken the human race in its grasp because of its zoonotic ability and spread across species. Monkeypox virus, an Orthopoxvirus of the Poxviridae family caused an alarming rate of infection among humans in the year 2022. The total number of cases reported so far is more than 50,000 and 16 deaths have been reported to the United Nations organization. Monkeypox was first detected among captive monkeys transported to Denmark from Africa. A larger pool of viruses is found to be reserved in rodents such as squirrels, and rats which are hunted down for food [35]. This disease has achieved global health importance after its outbreak in the USA in the year 2003 associated with infected pet prairie dogs. The dogs housed with rodents from Ghana, Africa became the primary source of infection and led to the infection in humans when they came in contact with the infected prairie dog. A cluster of infected individuals was reported in the summer of 2003 in the US Midwest and a larger outbreak of monkeypox was reported in Nigeria in 2017 [2]. After the first report of Monkeypox in May 2022, 92 laboratory-confirmed cases and 28 suspected cases were reported in non-endemic countries including Canada, Belgium, Australia, the Netherlands, Spain, Germany, France, Portugal, the United Kingdom, and the United States of America [36].

The variola virus (VARV) of the orthopox genus was considered to be the most dreaded one causing 300-500 million deaths across the globe, which was eradicated using live vaccinia virus (VACV) [37]. Monkeypox also belongs to this notorious group and is characterized by the formation of maculopapular lesions of about 2-5mm in diameter and spread in a centrifugal manner on a human body and rarely in a centripetal fashion. The lesions progress through various stage of formation namely, papular, vesicular, and pustular which

finally falls off the body leaving a dyspigmented scar. The disease is also known to affect the lymph nodes (lymphadenopathy) and is experienced with fever, chill, and body aches[1][37][38].

The mode of transmission of this disease from animals to humans is still elusive. Transmission through aerosols was demonstrated to prove the nosocomial outbreak in the Central African Republic. Indirect or direct contact with either dead or live animals is also considered a conduit of infection among humans. Exposure to bodily fluids such as saliva, respiratory exhalant, or exudates of lesions/scars is also a mode of transmission of the virus. The fecal matter of infected animals also harbours the virus and is responsible for infecting populations of an underdeveloped nation where people still sleep on the ground. The human-to-human transmission occurs through close contact, sharing contaminated objects, direct contact with lesions, and respiratory droplets. The degree of transmissibility of this disease i.e., the R0 value lie between 1.0 to 2.4 suggesting that an infected individual is capable of transmitting this disease to one or two more individuals, and in such a case it is advisable to follow quarantine protocols and maintain social distancing. Individuals with immunodeficiency especially those infected with HIV are at a higher risk of attaining the infection if not taken care of.

The diagnosis of this disease is done either by nucleic acid amplification technique for the estimation of DNA of the orthopoxvirus or specifically monkeypox virus. The quantification can also be assessed using real-time PCR assay and also by serological testing of IgG which becomes detectable post 8 days of the onset of symptoms. But it is advisable to not use antibody testing alone for the diagnosis of monkeypox[34].

The transmission of monkeypox can be controlled by following certain precautionary measures such as isolating infected individuals from the healthy ones and following quarantine procedures to control the spread. The human-to-human transmission can be reduced by avoiding close contact, physical touch, and sharing of things. The health co-workers that take care of patients should self-isolate them and cover any visible lesions to avoid transmission. A medical mask and usage of gloves are recommended while in close contact with an

infected person sanitization and following hygienic practices can also bring down the risk of transmission. Besides following sanitary and quarantine practices antiviral drugs used for the treatment of smallpox such as cidofovir, brincidofovir, and tecovirimat, and is used for the treatment of immunocompromised individuals infected with monkeypox. In addition to these, a vaccinia immune globulin approved by the US FDA can also be recruited for the treatment of severe cases. People vaccinated against smallpox may have some protection against monkeypox. The usage of US FDA-approved IMAVUNE is used for adults above 18 years at high risk of smallpox or monkeypox. The use of ACAM2000 licensed by the FDA, a smallpox vaccine is recommended for usage in Europe[34]. The non-availability of vaccines against monkeypox brings in a lot of opportunities for drug companies to design and synthesize vaccines specific to monkeypox. Vaccines provide protection not just to the virus but also prevent the spread of vector-borne illnesses. The development of a vaccine can be done using the conventional method which would involve the usage of protein, peptides, or inactivated/attenuated pathogens in combination with adjuvants. Unlike the conventional method, the immunoinformatic approach to developing a vaccine is considered to be cost-effective and takes a lesser time frame than the traditional approach[38][39].

The preliminary goal of our study is to adapt to the immunoinformatic approach of vaccine development and analyze and use the results obtained *in-silico* to develop a multiepitope vaccine for the treatment of monkeypox. The antigenicity and allergenicity of the vaccine developed via this method are taken into consideration for further studies. This approach also involves minimal risk to the individual because no live pathogen is used for the development of the vaccine. A 3D model of the vaccine is obtained by employing the fusion of B-cells and T-cells epitopes to build a new subunit of vaccine to activate and build innate, humoral, and cell-mediated immunity against the infection. The obtained multiepitope vaccine is also docked with the TLR4 receptor by MD simulation to verify the complex structure of the 3D-modelled vaccine obtained[39].

4.2. Materials and Methods:

4.2.1. Retrieval of antigenic proteins from the monkeypox virus genome:

The entire protein sequences of the Monkeypox virus were retrieved from the NCBI protein database in FASTA format (https://www.ncbi.nlm.nih.gov/protein/). These proteins were subjected to a web-based tool VaxiJen for analyzing their antigenicity[40]. The antigenic scores were obtained using the default parameters, and proteins showing an antigenic score >0.4 were carried forward for further investigation.

4.2.2. Prediction of MHC-I specific cytotoxic T-lymphocytes epitopes:

The prediction of CTL epitopes was analyzed using the NetCTL-1.2 server. The epitopes predicted by NetCTL-1.2 is a collective prediction of TAP transporter efficiency associated with antigen processing, proteasomal C-terminal cleavage, and MHC class-I binding resulting in the prediction of epitopes that are top-scoring and highly sensitive[40]. The selected proteins were subjected to an analysis against all the 14 available alleles (**HLA-A0l:01, HLA- A02:0l, HLA-A02:03, HLA-A02:06, HLA-A03:0l, HLA-All:0l, HLA-A24:02, HLA-A31:0l, HLA-A33:0l, HLA-B07:02, HLA-B35:0l, HLA-B51:01, HLA-B54:01, HLA-B58:01)** using default parameters and the epitopes showing the highest binding score and probable non-allergen were selected for the construction of a vaccine.

4.2.3. Prediction of MHC-II specific helper T-lymphocyte epitopes:

The NetMHCII Pan server was utilised to predict the 15-mer HTL epitopes showing higher binding affinity against the twenty MHCII alleles **DRBl_0l0l, DRB1_0301, DRB1_0401, DRB1_0405, DRB1_0701, DRB1_0802, DRB1_0901, DRB1_1101, DRB1_1201, DRB1_1302, DRB1_1501, HLA-DPA10103-DPB10101, HLA-DPA10201- DPB10201, HLA-DPA10301-DPB10301, HLA-DQA10401-DQB10401, HLA-DQA10501-DQB10501, HLA-DQA10102-DQB10202, HLA-DQA10402-DQB10402.** The epitope showing the highest binding score and probable non-allergen was chosen for vaccine construction.

4.2.4. B-cell epitope prediction:

To produce antibodies against a pathogen, the B-cells need to be activated by the specific antigens/epitopes. Therefore, the selected antigenic proteome of the Monkeypox virus was examined for the promiscuous linear B-cell epitopes by using the BCPRED server[41]. The BCPRED server takes the FASTA sequence as input and generates 20mer epitopes with the default specificity of 75% towards the B-cell receptors. The epitope with the highest score was selected for further multi-epitope vaccine construction.

4.2.5. Population coverage of the vaccine construct:

The world population coverage of the vaccine construct was determined on the basis of the frequency of occurrence for the respective HLA alleles. The population coverage tool of IEDB (http://tools.iedb.org/population/) that uses the allelfrequencies.net database was utilized to check the population coverage of the HLAs used in the analysis.

4.2.6. Designing of multi-epitope subunit vaccine sequence:

The Immunoinformatics approach is taken to predict BCL, CTL & HTL epitopes which are highly conserved, antigenic, non-toxic & non-allergenic and were used in vaccine construction. To create a multivalent vaccination design for MPXV, the epitopes were combined with various linkers and adjuvants. Antigenic nature is required in vaccines for the elicitation of elongated immune responses in the body. Linkers are required to make protein more stable as it helps in protein folding and in getting flexible conformation[42]. The Role of Adjuvant in Vaccines to boost immunogenicity and stimulate both innate and Adaptive. Two adjuvants were used, the first adjuvant was an agonist for TLR-4 Heparin-binding hemagglutinin (hbhA) which is a peptide of Mycobacterium tuberculosis and another one is TpD, the derivative of tetanus and diphtheria toxoid which elicit a humoral immune response on interaction with MHC class II alleles.

4.2.7. Evaluation of allergenicity, antigenicity, and Physiochemical properties of Multi-epitope vaccine construct:

We use three servers to analyze the allergenicity of the designed vaccine AllerTOP v.2.0 (https://www.ddg-pharmfac.net/AllerTOP/), Allergen FP v.1.0 (https://www.ddg-pharmfac.net/AllergenFP/) and AlgPred (https://webs.iiitd.edu.in/raghava/algpred/submission.html). AllerTOP and Allergen FP uses the data of known 2427 allergen and non-allergen of different species based on the K-nearest algorithm to predict allergenicity and AlgPred online server use a combination of tool (such as SVM, ARPs Blast, and MAST) to predict the allergenicity. For the prediction of antigenicity, we used two servers, Vaxijen v.2.0 server, with a threshold value of 04 another server was ANTIGENpro which can be accessed through (https://scratch.proteomics.ics.uci.edu/). The Epaxy ProtParam server (https://web.expasy.org/protparam/), is used to estimate the various physiochemical properties of designed vaccines such as Molecular weight, Isoelectric point (pI), Instability index, aliphatic index, In-vivo and In-vitro half-life and Grand average hydrophobicity value (GRAVY). The toxicity of the designed vaccine was analyzed through the ToxinPred server (http://crdd.osdd.net/raghava/toxinpred/). The solubility of the protein sequence was also checked using the SOLpro server which is based on SVM based technique to predict with an average of 75% accuracy[43].

4.2.8. Prediction of Secondary and Tertiary structure of the vaccine construct and Refinement:

The SOPMA server tool was used in the prediction of the secondary structure (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa sopma.html) vaccine construct. SOPMA stands for (Self Optimized Prediction Method with Alignment) and it uses a method based on a homology model. The I-Tasser server (https://zhanggroup.org/I-TASSER/) was used to predict the tertiary structure of the multi-epitope vaccine construct. I-Tasser is based on confidence score (C-Score). A higher C-Score indicates that the developed model is more accurate. Top-5 models get selected based on C-Score, Ramachandran plot of these 5 models gets constructed for a further selection of

a model having a higher percent residue in the favorable region. Further, the modeled 3-D structure was refined by using the Galaxy Refine server (https://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE). This server is based on the ab initio modeling method which corrects loop or terminal regions of the predicted 3-D model. Further, validation of refined structure was done in the PROCHECK server using SAVES 6.0 tool (https://saves.mbi.ucla.edu/) [44].

4.2.9. Molecular dynamic simulation of the vaccine construct:

The vaccine construct was further checked for its stability by molecular dynamics (MD) simulation using the pmemd.cuda module of the AMBER18 package. The ff14SB force field was used to describe the amino acid residues, with the tLEaP module utilized for adding the missing hydrogen residues and solvation. The solvation of the vaccine construct was done in a TIP3P octahedron periodic box of 10 Å distance from the protein's surface. Additional Sodium (Na+) and Chloride ions (Cl-) were added for system neutralization and to make it in conformity with the physiological salt concentration of 150 mM. The SHAKE algorithm was employed to restrict the motion of bond lengths that included hydrogen atoms. The long-range electrostatic interactions were computed from the Particle Mesh Ewald (PME) method. The system was minimized in two steps. In the first step, restrained energy minimization was carried out using 5000 steepest descent steps followed by another 5000 conjugate gradient steps. The second step involved unrestrained minimization with 100 steepest descent followed by 900 conjugant gradient steps. The system was heated from 0 K to 300 K in multiple stages and the temperature was maintained using the Langevin thermostat. Following a 1 ns equilibrium in the NVT ensemble, a 200 ns NPT production run was carried out where the Berendsen Barosat was used to maintain a constant pressure of 1.0 bar. During the production simulation, the Cartesian coordinates were saved at an interval of 10 ps. In total, 20,000 snapshots were collected. The RMSD analysis was conducted using the Cpptraj module of AMBER18. The structure corresponds to a stable snapshot at appx. 199.52 ns was taken for further analysis. Structure

validation was done using the SAVES 6.0 tool and the Ramachandran plot was observed.

4.2.10. Molecular docking of multi-epitope vaccine construct with TLR4 and binding affinity prediction

The vaccine construct had a TLR4 adjuvant at its N terminal end. Protein-protein docking was carried out in the ClusPro web server (https://cluspro.bu.edu/home.php) to check the vaccine's interaction with TLR4. We used the PDB ID 3FXI as the initial structure of TLR4 for our studies. During submission in the ClusPro server, TLR4 was considered as the receptor while the vaccine construct was considered as the ligand. The docking calculation in ClusPro is done with the following steps that include a rigid body-based docking by FTT-based global sampling on a grid using PIPER, followed by an RMSD-based clustering to find highly populated clusters of low energy conformations, and further refinement by CHARMM minimization to remove steric clashes. A total of 30 docked complexes were generated in ClusPro based on the weighted score for the energy of the complex between the receptor and ligand. The ClusPro weighted scores are calculated as:

$$\mathbf{E=0.40E_{rep}+-0.40E_{att}+600E_{elec}+1.00E_{DARS}}$$

"where E_{rep} and E_{attr} denote the repulsive and attractive contributions to the van der Waals interaction energy, and E_{elec} is an electrostatic energy term. E_{DARS} is a pairwise structure-based potential constructed by the Decoys as the Reference State (DARS) approach, and it primarily represents desolvation contributions, i.e., the free energy change due to the removal of the water molecules from the interface."

Further, we predicted the binding affinity and dissociation constant of the docked complexes using the Prodigy haddock webserver (https://wenmr.science.uu.nl/prodigy/). The binding affinity ΔG is calculated as:

$$\mathbf{\Delta G_{predicted} = -0.09459\ ICs_{charged/charged} -0.10007\ ICs_{charged/apolar} +0.19577\ ICs_{polar/polar} -0.22671\ ICs_{polar/apolar} +0.18681\ \%NIS_{apolar} +0.3810\ \%NIS_{charged} -15.9433}$$

Where "$Cs_{xxx/yyy}$ is the number of Interfacial Contacts found at the interface between Interactor1 and Interactor2 classified according to the polar/apolar/charged nature of the interacting residues (i.e. $ICs_{charged/apolar}$ is the number of ICs between charged and apolar residues). Two residues are defined as in contact if any of their heavy atoms is within a distance of 5.5 Å" (Xue et al., 2016)

Based on the predicted ΔG the dissociation constant is further calculated by the formula:

$$\mathbf{\Delta G = RT\ lnK_d}$$

Where R, T, and ΔG respectively denote the ideal gas constant (in kcal K-1mol-1), temperature (in K), and the predicted free energy (in kcal/mol). The temperature for this experiment was set at 300 K (27.0 ℃).

4.2.11. Molecular dynamic simulation to check the Stability of the multi-epitope vaccine constructs and vaccine-TLR4 complex

To assess the structural stability of the vaccine construct while bound to TLR4, a molecular dynamics (MD) simulation of 250 ns was carried out using the same protocol as previously described. The aim was to analyze the complex system's overall stability throughout an extended period. All analyses were performed using the Cpptraj module of AMBER18. The most stable complex structure corresponding to the most probable RMSD value obtained from the production simulation. PDBSum webserver (http://www.ebi.ac.uk/thornton-srv/databases/pdbsum/Generate.html) was further used to identify the crucial residues responsible for the interaction between the molecules.

4.2.12. MM-PBSA and critical residue analysis of the multi-epitope vaccine construct—TLR4 complex

The molecular mechanics Poisson Boltzmann surface area (MM-PBSA) method was employed to compute the binding free energy of the Vaccine -TLR4

complex formation. The binding free energy was calculated by implementing the following equations:-

$$\Delta G_{bind} = \Delta H\text{-}T\Delta S\ \Delta E_{internal} + \Delta G_{solv} - T\Delta S$$

$$\Delta E_{internal} = \Delta E_{covalent} + \Delta E_{elec} + \Delta E_{vdW}$$

$$\Delta Gsolv = \Delta G_{pol} + \Delta G_{np}$$

where the terms $\Delta E_{internal}$, ΔG_{solv}, and $T\Delta S$ are used to indicate the total internal energy, solvation-free energy, and conformational entropy, respectively (Miller et al., 2012). The internal energy is further divided into energies of $\Delta E_{covalent}$ (bond, dihedral, and angle), ΔE_{elec} (electrostatic), and ΔE_{vdW} (van der Waals). The desolvation-free energy is composed of polar (ΔG_{pol}) and non-polar (ΔG_{np}) free energies. A total of 2000 structural frames at a frequency of 30ps from the last 60ns trajectory was used for the enthalpy (ΔH) calculations. We could not perform the estimation of conformational entropy (ΔS) due to its high computational cost. To identify the key residues that stabilized the TLR4-vaccine complex, MM-PBSA per-residue decomposition was conducted.

4.2.13. cDNA construct, codon optimization, and in silico cloning of designed vaccine:

The constructed vaccine cDNA was generated by importing their amino acid sequence in the Reverse translation tool of the EXPASY server. The main sequence of the vaccine constructs was evaluated for expression rate in a suitable expression vector (host *Escherichia coli* strain K12) by using the Java Codon Adaptation Tool (JCAT) (http://www.jcat.de/). The protein expression potency is checked through the JCAT result which consists of the Codon Adaptation Index (CAI) whose ideal value is 1.0 and GC content percent whose ideal range is between 30% and 70% of the total GC amount. The NEB Cutter V 2.0 () was used to search for restriction sites in the optimized cDNA and these sites were identified to be appended. For *in-silico,* cloning, the SnapGene software was used where the final gene construct was cloned in the pET28a (+) expression vector.

4.2.14. Immune simulation:

We adopted the C-IMMSIM v10.1 tool to submit the developed vaccine for an assessment of its potential immunological effectiveness (C-IMMSIM Online server available at http://kraken.iac.rm.cnr.it/C-IMMSIM)[44]. The C-IMMSIM server uses position specific scoring matrix (PSSM) method to identify epitopes and stimulate the immunological response. The injection of the vaccine did not contain LPS, and the simulation's volume was set to 10 with a 12,345 random seed. There were 1000 simulation steps in all, with injection time steps of 1, 84, and 168. The other simulation parameters were left at their default values. The gap between each dose was 28 days as suggested in different literature[45].

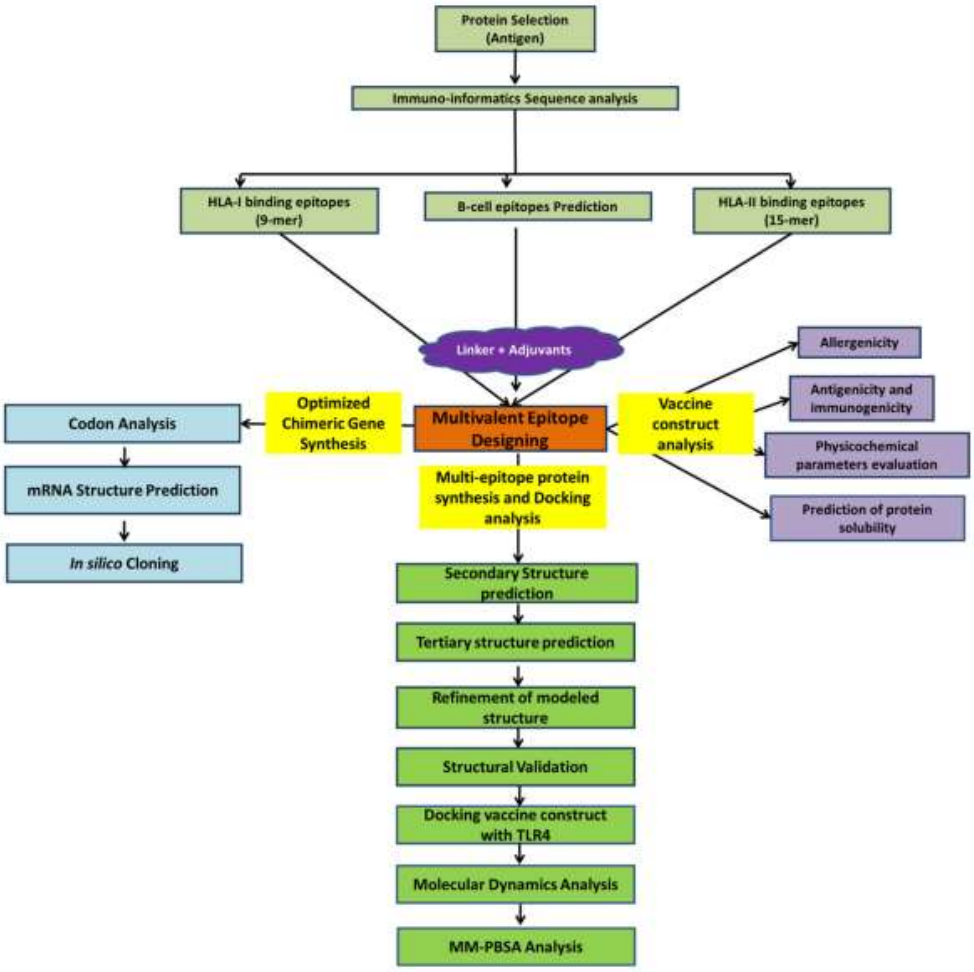


Figure 4.1. Flowchart of the immuno-informatic strategy used for the MPXV vaccine construct analysis.

4.3. Results and Discussion:

4.3.1. Selection of proteins for the epitope designing and antigenicity analysis for the construction of the vaccine.

The essential proteins were retrieved from the NCBI database in FASTA format and were analyzed for antigenicity using the web-based tool VaxiJen. An effective vaccine can only be constructed using proteins showing adequate immunogenic response. For viruses, the threshold was set to 0.4, proteins showing an antigenicity score >0.4 were selected for the construction of a vaccine, and those reflecting antigenicity score <0.4 were considered non-antigenic. The list of proteins showing antigenic scores more than the threshold is mentioned in (Table 4.1).

Table 4.1. Prediction of antigenicity of Monkeypox virus protein from Vaxigen server.

S.No.	Gene Name	Protein Name	NCBI Protein ID	VaxiJen Score
1	**A35R**	EEV glycoprotein (1)	NP_536572.1	**0.5449**
2	**A36R**	EEV glycoprotein (2)	NP_536573.1	**0.4438**
3	**A38R**	IEV transmembrane phosphoprotein	NP_536575.1	**0.4038**
4	**B6R**	EEV type-I membrane glycoprotein	NP_536594.1	**0.5773**
5	**M1R**	IMV membrane protein L1R	NP_536507.1	**0.6112**

6	**E8L**	Carbonic anhydrase	NP_536532.1	**0.5076**
7	**A5L**	A5L protein-like	NP_536542.1	**0.4150**
8	**H3L**	IMV heparin binding surface protein	NP_536520.1	**0.4481**
9	**A21L**	IMV membrane protein A2	NP_536558.1	**0.4196**
10	**A32L**	IMV membrane protein A30	NP_536569.1	**0.7336**

4.3.2. Prediction of MHC-1 specific CTL epitopes:

CTLS are responsible for directly killing other cells and play a pivotal role in host defense against infections caused by viruses and pathogens that replicate intracellularly in the cytoplasm. NetCTL server was used to predict the CTL epitopes from the selected MPXV proteins. The method amalgamates the prediction of peptide MHC-I binding, Proteasomal C-terminal cleavage, and TAP transport efficiency. This server utilizes artificial neural networks that help in predicting a 9-mer showing a high propensity to proteasomal cleavage at its C-terminal ends and high affinity to MHC-I molecules. In the present study we used 14 different alleles of MHC-I namely: **HLA-A0l:01, HLA- A02:0l, HLA-A02:03, HLA-A02:06, HLA-A03:0l, HLA-All:0l, HLA-A24:02, HLA-A31:0l, HLA-A33:0l, HLA-B07:02, HLA-B35:0l, HLA-B51:01, HLA-B54:01, HLA-B58:01**. To obtain high epitope-HLA interaction specificity of more than 97%, the NetCTLs server's score was set as 0.4. the selected epitopes were further analyzed as one epitope per allele was selected having higher binding affinity and non-allergen (Table 4.2).

Table 4.2. List of CTL epitopes

SL. NO.	Cytotoxic T cell epitope	Start (aa)	End (aa)	Binding Affinity
1.	VSDYVSELY	78	86	0.996
2.	SLSAYIIRV	184	192	0.965736
3.	TLKDLMSSV	229	237	0.984114
4.	STLDYFTYL	162	170	0.902861
5.	KIKSPNSSK	53	61	0.971561
6.	RVLFSIFYK	91	99	0.939823
7.	RYPGVMYTF	267	275	0.997417
8.	KSQESYYQR	42	50	0.966961
9.	ESYYQRQLR	45	53	0.810342
10.	VPAMFTAAL	87	95	0.935135
11.	QPVKEKYSF	141	148	0.942536
12.	VPLITVTVV	12	20	0.951341
13.	TPLQPPIVA	187	195	0.976295
14.	LANKENVHW	220	228	0.995694

4.3.3. Prediction of MHC-II specific helper T-lymphocyte (HTL) epitopes:

HTL epitopes are crucial to both humoral and cellular immune response owing to their participation in Class II MHC activation. MHC-II specific HTL epitope (15-mers) were predicted by the IEDB tool. To cover the maximum population throughout the world, a set of 1 MHC-II molecules namely: **DRBl_0l0l, DRB1_0301, DRB1_0401, DRB1_0405, DRB1_0701, DRB1_0802,**

DRB1_0901, DRB1_1101, DRB1_1201, DRB1_1302, DRB1_1501, HLA-DPA10103-DPB10101, HLA-DPA10201- DPB10201, HLA-DPA10301-DPB10301, HLA-DQA10401-DQB10401, HLA-DQA10501-DQB10501, HLA-DQA10102-DQB10202, HLA-DQA10402-DQB10402. To obtain high epitope-HLA interaction specificity of more than 97%, the NetCTLs server's score was set as 0.4. the selected epitopes were further analyzed as one epitope per allele was selected having higher binding affinity and non-allergen (Table 4.3).

Table 4.3. List of HTL epitopes

SL. NO.	T Helper cell epitope	Start (aa)	End (aa)	Binding Affinity
1.	DSGYHSLDPNAVCET	49	63	0.972513
2.	AKYVEHDPRLVAEHR	234	248	0.969403
3.	PNFWSRIGTVAAKRY	19	33	0.850371
4.	LTNFKQLNSTTDSEA	126	140	0.964813
5.	DIDVQYTSTLSVVHE	21	35	0.912427
6.	YKDYWVSLKKTNDKW	98	113	0.838941
7.	GEIIRAATTSPVREN	235	249	0.945093
8.	LTAILFLMSQRYSRE	287	301	0.751698
9.	FKIFNINKKSKKNSK	42	56	0.805494
10.	RNELTIYQNTTVVIN	134	148	0.970336
11.	WNKKKYSSYEEAKKH	96	110	0.515812
12.	QEKRDVVIVNDDPDH	44	58	0.7849

4.3.4. Linear B-cell epitopes prediction:

B-lymphocytes are a critical component of mammals' adaptive immune response responsible for humoral immunity. These are responsible for generating diversified but pathogen-specific antibodies resulting in antigen neutralization and viral load clearance. To include the epitopes having the propensity of activating the B-cells in the vaccine construct, the antigenic proteome was analyzed by using the ABCpred server. The best epitope from each antigenic protein with the highest score was used for the vaccine construction (Table 4.4.).

Table 4.4. List of B-cell epitopes

SL. NO.	PROTEIN NAME	B cell epitope	Start (aa)	Antigenic score
1.	EEV glycoprotein (1)	VIRSMMTPENDEEQTS	23	0.92
2.	EEV glycoprotein (2)	NGWIQYDKHCYLDTNI	75	0.91
3.	IEV transmembrane phosphoprotein	TSICSKSNPFIAELNN	216	0.94
4.	EEV type-I membrane glycoprotein	CDSGYHSLDPNAVCET	86	0.94
5.	IMV membrane protein L1R	MVMEMRANEPRTEINL	4	0.9
6.	Carbonic anhydrase	GEIIRAATTSPVRENY	263	0.92
7.	A5L protein-like	QPTIHITPQPVPTPTP	98	0.91
8.	IMV heparin binding surface protein	HEHINDQKFDDVKDNE	67	0.97
9.	IMV membrane protein A21	AISEKMRRERAAYVNY	51	0.93
10.	IMV membrane protein A30	MVMEMRANEPRTEINA	4	0.91

4.3.5. Population coverage analysis:

The population coverage worldwide was predicted using the IEDB server which is based on the interaction of selected epitopes with a number of HLA alleles as the expression of HLA alleles varies in different countries which are affected by the illness or diseases like the monkeypox virus. The worldwide population coverage of the MHC-I epitope was 91.22% while MHC-II coverage was 99.67%. Combinedly, the MHC-I and MHC-II show 99.97% coverage of the worldwide population *(Figure 4.2).*

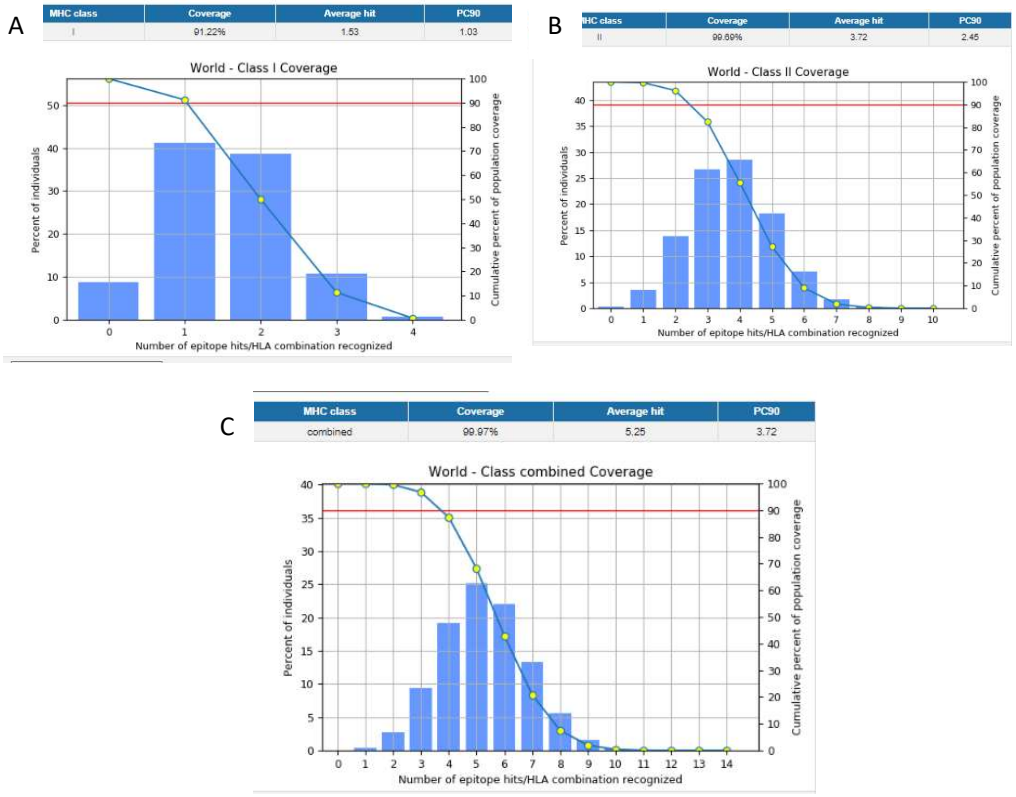


Figure 4.2 Predicted epitope population coverage analysis (A) MHC-I allele (B) MHC-II allele (C) combined MHC-I and MHC-II allele.

4.3.6. Construction of multi-epitope subunit vaccine:

The construction of a multi-epitope vaccine was done by merging CTL (14), HTL (13), and BCL (9) epitopes with different adjuvants and linkers. MHC-I binding epitopes were joined by an AAY linker which helps in binding of

antigen epitopes to the TAP transporter that leads to enhancement in epitope expression. KK linkers were used in joining the B-cell epitope. HTL epitopes get stimulated by GPGPG linkers to increase immunogenicity. To further increase the immunogenicity and antigenicity of the constructed vaccine two adjuvants were added, first one was added to the N-terminal side of constructed vaccine via EAAAK linker. The first adjuvant was similar to TLRs, as they upon encounter with antigen trigger the signaling cascade for an adaptive immune response within the host by increasing the production of Interferon-Beta. Heparin-binding Hemagglutinin (hbhA) has a similar response and activity as TLR4 [46], when added to the constructed vaccine it increases the interferon beta-mediated immune response and reported in cancer vaccines (https://doi.org/10.1158/0008-5472.CAN-10-3487)[49]. Another adjuvant TpD was added between CTL and HTL epitopes for enhancement of adaptive humoral immune response. The pan-HLA DR binding epitope (PADRE) can attach to a variety of mouse and human MHC-II alleles to produce CD4+ helper cell response which further increases vaccine immunogenicity. The CTL, HTL, and B cell epitopes, along with hbhA-TLR4, TpD adjuvant, and all linkers, form a 736 amino acids long multi-epitope vaccine construct (*Figure 4.3)*

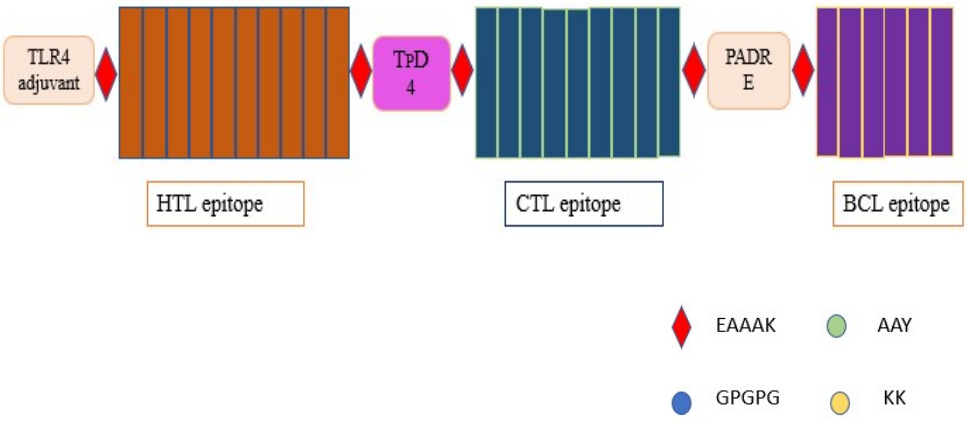


Figure 4.3 Linear representation of the construct multiepitope vaccine of MPXV

4.3.7. Multi-epitope subunit vaccine construct allergenicity, immunogenicity, and physiochemical analysis

A constructed vaccine to be successful and safe must have the highest level of immunogenicity and which does not cause any allergies. Therefore, we verify the allergenicity on three different servers for authenticity. The AllerTOP and Allergen FP servers predicted the allergenicity of the constructed vaccine as non-allergen based on the K-nearest algorithm. AlgPred also predicted it as a non-allergen based on a score = -0.21 which signifies non-allergenicity. The immunogenicity of the vaccine was checked by Vaxijen and ANTIGENpro server. For the virus model, the Vaxijen server has a threshold value of 0.4, and any value above 0.4 is counted as antigenic, the predicted score was 0.4889 for a constructed vaccine. ANTIGENpro server gives the score on the basis of microarray data, where > 0.5 scores is taken as good antigen and a score of vaccine construct was 0.9322. The web-based tool ProtParam was implemented to derive the physiochemical properties of the constructed vaccine. The chemical formula of the constructed vaccine is $\mathbf{C_{4093}H_{6502}N_{1134}O_{1231}S_{20}}$ with a molecular weight of 91.9kDa. The molecular weight of constructed vaccine is higher than the ideal mol. wt of vaccines. As reported higher mol. Wt higher the chances to come in contact with humoral antibodies to elicit the immune response. The isoelectric point of the construct is 9.4 reflecting that it is basic. The total number of amino acids in the construct is 840, which is soluble, and non-toxic. The stability of the vaccine is predicted by the instability index where the score should be <40. The instability index is 35.35 owing to the stability of the construct. The half-life of the vaccine construct is observed to be >30 h in mammalian reticulocytes, >20 h in yeast, and > 10 h in *Escherichia coli*. The predicted result signifies that if a host is subjected to a prolonged period then it might strengthen the immune system. The GRAVY (grand average of hydropathy) scores signify the hydrophilic in nature if positive or hydrophobic if negative. The GRAVY score of the vaccine construct was -0.548 which indicates the hydrophobic or non-polar nature of the vaccine. The SolPro data depicts that the construct is soluble with a probability of 93%. While toxicity analysis on the ToxinPred server indicates the non-toxic behavior of the construct (*Table 4.5).*

Table 4.5. The obtained antigenicity, allergenicity and physiochemical properties.

Sl. No.	Features	Assessment	Remarks
1.	Antigenicity	• VaxiJen Server 0.4889 • AntigenPro Server Predicted Probability of Antigenicity: 0.930220	Probable ANTIGEN
2.	Allergenicity	Confirmed by -: • AllerTop v.2.0 Tool • AllergenFP v.1.0 Tool • AlgPred v.2.0 Tool	Probable non-allergen Non allergen Score = − 0.21 Probable non-allergen
3.	Solubility	• SolPro Tool : Soluble with probability = 0.976850	Soluble
4.	Toxicity	• ToxinPred 2.0 Tool	Non-Toxin
5.	Number of amino acids	840	Suitable
6.	Molecular weight	91.9 kDa	Average
7.	Theoretical Isoelectric point (pI)	9.40	Basic
9.	Chemical Formula	$C_{4093}H_{6502}N_{1134}O_{1231}S_{20}$	-
11.	Estimated half-life	30 h (mammalian reticulocytes, in vitro). >20 h (yeast, in vivo). >10 h (Escherichia coli, in vivo).	-
12.	Instability index	35.35	Stable
13.	Aliphatic index	72.63	Thermostable
14.	Grand average of hydropathicity (GRAVY)	- 0.548	Hydrophilic

4.3.8. Structure prediction and validation of the multi-epitope subunit vaccine construct:

The secondary structure prediction of the vaccine construct was analyzed by using the SOPMA server which helps in estimating the interaction of the vaccine construct with immunological receptors like TLRs. Its analysis depicts the result as ~46% alpha helix, ~16% beta-sheet, ~33% random coil and ~ 5% beta turn are present in the vaccine construct (*Figure 4.4*).

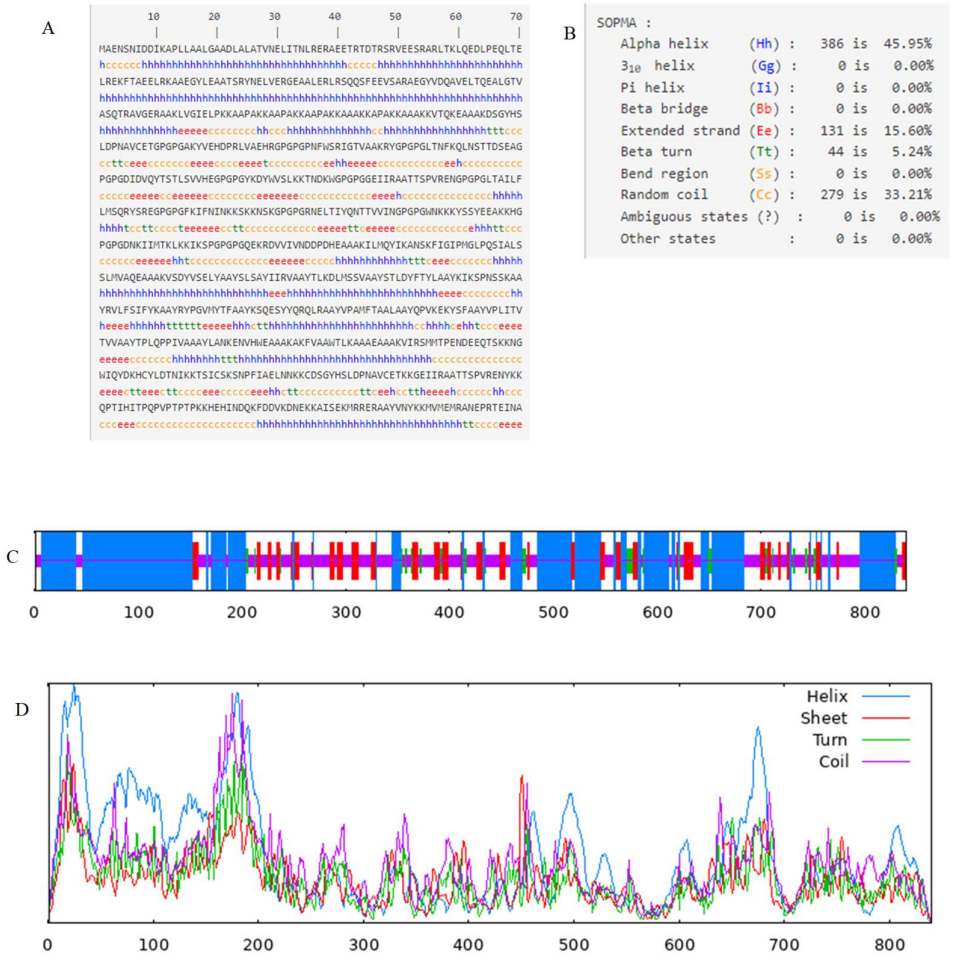


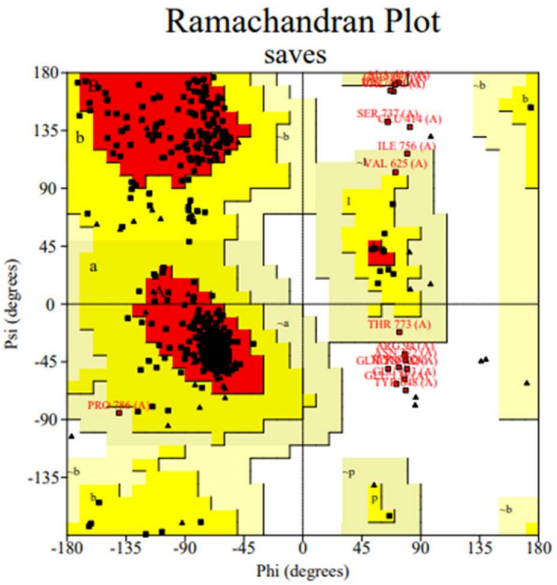
Figure 4.4.Secondary structure predictions of MPXV vaccine construct. A. The sequence of the vaccine construct along with the predicted secondary structure. B. The overall percentage of various secondary structures as predicted by SOPMA server. C Pictorial representations of various secondary structures. D. Propensity of occurrence of various secondary structures according to the residues in the vaccine construct. The secondary structure at a particular residue was predicted by the height of the peak

I-TASSER predicts the complete 3D structure of the given peptide sequence based on the multi-template-based threading method and provides several models with a rank of C-Score. C-Score ranges between -5 and 2 and a higher C-Score of the model depicts the higher confidence of a model. The first model having a higher C-Score of -1.32 among all the constructed models gets selected. And, further Validation of the selected model is done through Ramachandran plot analysis by using the PROCHECK server which deduced that 80.9% of residues of the predicted structure lay in the most favorable region and 15.2% of residues lie in the additionally allowed region. The modeled structure is further refined by using the GalaxyRefine server. It generates five structure models *(Figure 4.5).* Ramachandran plot was constructed of all the models and the model having higher percent residues in the allowed region is selected. The best model having 91.3% residues in the most favorable region along with 6.3% in the additionally allowed region and only 0.7% in the generously allowed region (Figure 4.6) generated by GalaxyRefine with GDT-HA of 0.93, RMSD 0.468, MolProbity 2.018 and clash score of 12.4 was selected as the best vaccine construct and used for further Molecular dynamics simulation analysis.

Structure Information

Model	GDT-HA	RMSD	MolProbity	Clash score	Poor rotamers	Rama favored
Initial	1.0000	0.000	3.054	7.2	16.2	79.6
MODEL 1	0.9315	0.468	2.018	12.4	0.3	93.8
MODEL 2	0.9310	0.471	2.037	12.5	0.6	93.4
MODEL 3	0.9423	0.460	2.025	12.7	0.6	93.8
MODEL 4	0.9405	0.452	1.975	10.8	0.4	93.6
MODEL 5	0.9342	0.474	2.029	12.8	0.6	93.8

Figure 4.5. Refined models from Galaxy Refine server.



Amino acid position	Residue number	Percentage (%)
Residues in most favored regions	665	91.3%
Residues in additional allowed regions	46	6.3%
Residues in generously allowed regions	5	0.7%
Residues in disallowed regions	12	1.6%
Total	728	100%

Figure 4.6. Tertiary structure validation of multi-epitope vaccine constructs Ramachandran plot after refinement showing allowed and disallowed region of amino acids.

4.3.9. Structural stability of the modeled vaccine construct

Following the molecular dynamics simulation, the trajectory was analyzed to gain insight into the structural stability of the vaccine construct. Analysis of the RMSD of the MPV vaccine construct revelated that the system became stable at ~125 ns and remained stable till 200 ns depicting the stability of the vaccine construct (Figure 4.7A). The kinetic, potential, and total energy of the system remained stable throughout the simulation indicating the structural stability of the vaccine construct. The stable structure around appx 199.52 was taken for the subsequent analyses.

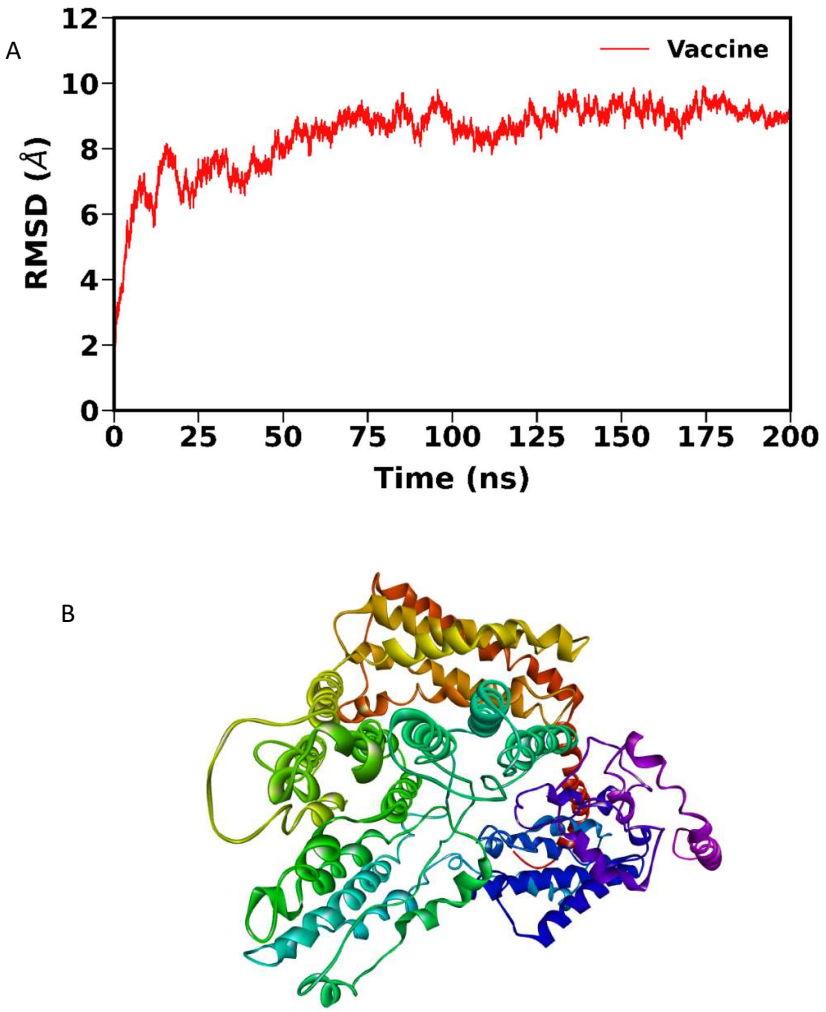


Figure 4.7. (A) RMSD Plot and (B) image of the final vaccine construct

4.3.10. Multi-epitope subunit vaccine constructs interaction with TLR4

A total of 30 docked structures were generated in ClusPro. Based on the weighted score for the energy, the top 10 models were further analyzed in Prodigy to predict the binding affinity(ΔG) and the dissociation constant (Kd) (*Table 4.6*). The model with the highest predicted binding affinity of -18.9 kcal/mol and the lowest ClusPro weighted score of -1316.7 was used for further analysis [47].

Table 4.6. ClusPro scores, binding affinity (ΔG), Dissociation constant (Kd) of top 10 models.

Cluster	Members	Representative	Weighted Score	ΔG	Kd
0	62	Center	-871.1	-16.8	5.90E-13
0	62	Lowest Energy	-1315.7		
1	41	Center	-887	-16.2	1.60E-12
1	41	Lowest Energy	-1316.7		
2	37	Center	-1147.1	-18.9	1.60E-14
2	37	Lowest Energy	-1147.1		
3	34	Center	-1091.3	-17	4.10E-13
3	34	Lowest Energy	-1092.5		
4	24	Center	-936.5	-17.3	2.60E-13
4	24	Lowest Energy	-979.2		
5	20	Center	-903.5	-16.4	1.20E-12
5	20	Lowest Energy	-1096.1		
6	20	Center	-1002.8	-16.8	5.80E-13
6	20	Lowest Energy	-1074.8		
7	19	Center	-894.6	-14.8	1.60E-11
7	19	Lowest Energy	-1008.9		
8	18	Center	-885.5	-15	1.20E-11
8	18	Lowest Energy	-941.5		
9	18	Center	-890.8	-16.1	2.00E-12
9	18	Lowest Energy	-992.8		

4.3.11. Structural stability analysis of the multi-epitope subunit vaccine construct complexed with TLR4 receptor

To assess the structural stability of the TLR4-vaccine complex, molecular dynamics simulation of 250 ns was carried out and the trajectory was analyzed for root-mean-squared deviation (RMSD), root-mean-square fluctuations (RMSFs), radius of gyration (Rg), and solvent accessible surface area (SASA). It is evident from Figure 4.9. that the TLR4-vaccine complex, as well as the vaccine construct, attained stability at around 100 ns (*Figure 4.9. A)*. Further analysis of the binding region RMSD revealed the stability of the interacting region of the TLR4 and the vaccine construct (*Figure 4.9. B)*. Next, we computed the root-mean-square fluctuations (RMSFs) of the complex residues only accounting for the Cα atoms, as shown in (*Figure 4.9. C)*. The Figure clearly depicts that the vaccine construct exhibited a higher level of residual

fluctuation as compared to TLR4. Further, the RMSF analysis indicates that the interface area of both the vaccine and TLR4 demonstrated relatively lower fluctuation, suggesting the overall stability of this region.

The Radius Gyration (Rg), which indicates the compactness of a system, was also stable throughout the simulation course (*Figure 4.9. E)* with a mean Rg of 50.09 (0.002) Å, reflecting that the complex maintained its overall compactness throughout the 250 ns simulation. The Solvent Accessible Surface Area (SASA) profile analysis of the vaccine—TLR4 complex showed an average value of 919.75 (0.06), with an overall stable time evolution observed in (*Figure 4.9. F)*, suggesting the stable nature of the TLR4-Vaccine interaction. The results of the MD simulation analysis indicate that the multi-epitope vaccine construct can interact stably with TLR4 and potentially trigger a robust immune response against the monkeypox virus.

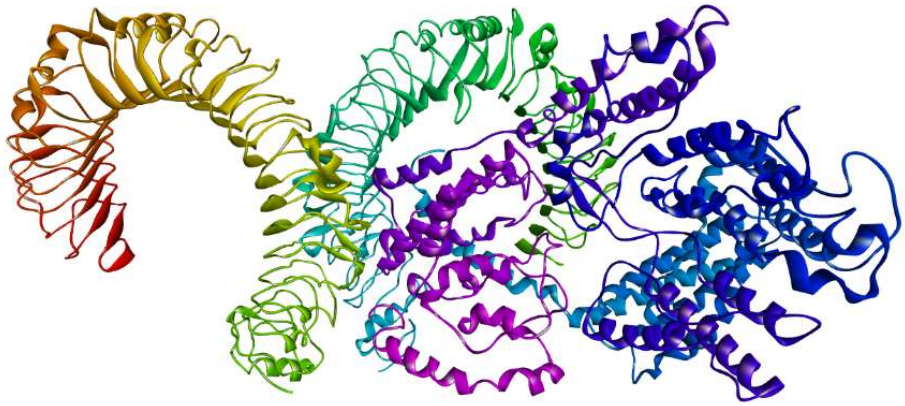
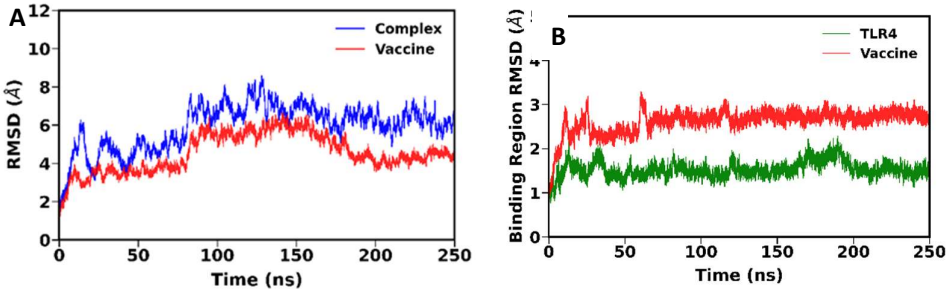


Figure 4.8. Image of TLR4- Vaccine docked complex.



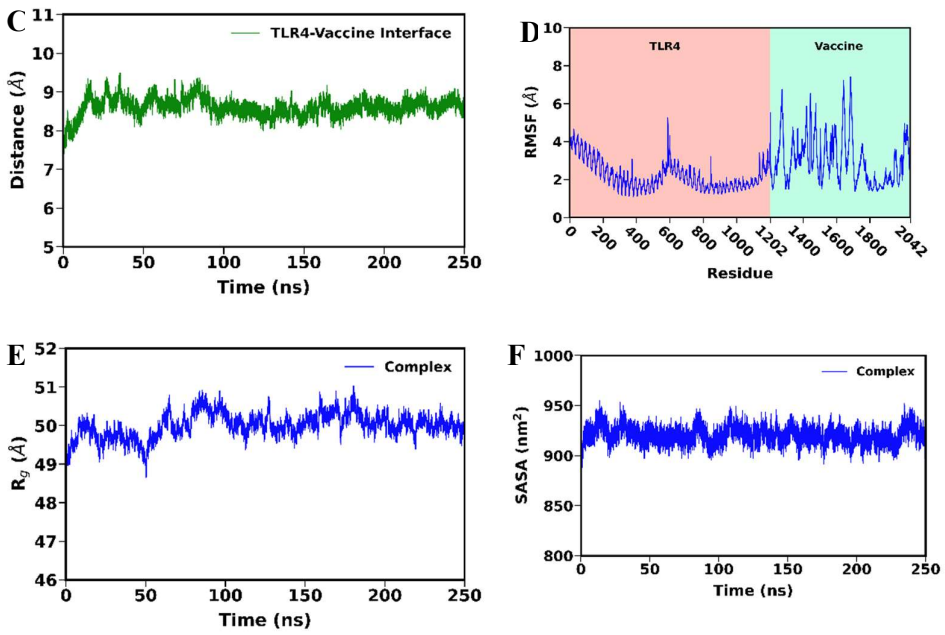


Figure 4.9. (A) RMSD plot of vaccine and docked complex (B) RMSD plot of binding region of TLR4-Vaccine construct, (C) Distance plot of TLR4-vaccine binding interface, (D) RMSF plot of TLR4 and vaccine, (E) Rg plot of docked complex, and (F) SASA plot of docked complex.

4.3.12. Free energy calculations and energy decompositions using MM-PBSA reveals the stability and high affinity of vaccine construct to the immune receptor

The molecular mechanics Poisson Boltzmann surface area (MM-PBSA) method was utilized to determine the binding free energy between the multi-epitope vaccine construct and TLR4, with the aim of examining the underlying energetics of their interaction. The binding free energy (ΔG_{Bind}) was calculated to be -241.84 kcal/mol.

Table 4.7. Calculated energies of TLR4-Vaccine complex

System	ΔE_{vdW}	ΔE_{elec}	ΔG_{pol}	ΔG_{np}	ΔG_{bind}
TLR-Vaccine Complex	-286.00 (0.24)	-6171.46 (2.08)	6247.21 (2.02)	-31.59 (0.02)	-241.84 (0.43)

From **Table 4.7**, we can interpret that the intermolecular electrostatic (ΔE_{elec} = -6171.46 kcal/mol) and van der Waals (ΔE_{vdW} = -286 kcal/mol) interactions, as well as non-polar solvation free energy ($\Delta G_{non\text{-}polar}$ = -31.59 kcal/mol), were crucial for the binding of the vaccine with TLR4. On the other hand, polar solvation-free energy (ΔG_{polar} = 6247.21 kcal/mol) disfavored the formation of vaccine/TLR4 complex. The net polar component of the binding free energy (elec + pol = 75.75 kcal/mol) disfavored the binding, which was counterbalanced by the van der Waals and the non-polar interaction. Overall, the van der Waal interactions were the most prominent contributor towards the binding of the molecules. Further, the MM-PBSA pairwise decomposition was utilized to identify the key residues involved in stabilizing the vaccine-construct-TLR4 complex system. Residues with energy contribution ≤ 1.5 kcal/mol are listed in (Annexure 1A). The most prominent contribution from the vaccine towards the binding was demonstrated by the residues Thr1783, Ser1939, Ile1721, Lys1820, Met1781, Leu1227, Tyr1720, Arg1723, Lys1935, Pro1804, Tyr1615, Lys1854, Tyr1782, Asn1206, among others (see table 4.8).

Table 4.8. Residues with energy contribution ≤ 2.0 kcal/mol are listed

Vaccine	**Energy (kcal/mol)**	**TLR4**	**Energy (kcal/mol)**
Thr1783	-4.58	Arg434	-6.98
Ser1939	-4.52	Gln458	-5.47
Ile1721	-4.49	Asn712	-5.13
Lys1820	-4.30	Lys534	-4.84
Met1781	-4.23	Phe983	-3.04
Leu1227	-3.78	Val605	-2.78
Tyr1720	-3.77	Pro688	-2.70
Arg1723	-3.60	Arg957	-2.65
Lys1935	-3.47	Pro487	-2.44
Pro1804	-3.36	Hid1104	-2.40
Tyr1615	-3.30	Leu618	-2.33
Lys1854	-3.24	Gln666	-2.19
Tyr1782	-3.10	Phe461	-2.10
Asn1206	-3.06	Glu413	-2.08

4.3.13. Hydrogen bonds and hydrophobic interactions

Hydrogen bonds are essential for the overall structural stability of a protein-protein complex. From the simulation trajectory, the time evolution of the number of hydrogen bonds formed between the vaccine and TLR4 was analyzed. The corresponding H-bond parameters are included in **Table 4.9**.

Table 4.9. Hydrogen bond parameters for corresponding residues

Acceptor	Donor	Dist.	Angle	Occupancy (%)
Asn1853@O	Gln458@NE2	2.84	160.91	70.94
Ser1718@O	Gln666@NE2	2.85	160.44	69.82
Met1781@O	Ser759@OG	2.75	156.00	52.78
Glu1855@OE2	Arg434@NH2	2.82	153.15	41.99
Glu1855@OE1	Arg434@NH2	2.82	151.95	41.17
Pro1804@O	Hid1104@ND1	2.84	159.21	40.92
Glu1855@OE1	Arg434@NH1	2.78	154.27	38.41
Glu1855@OE2	Glu459@N	2.84	160.57	37.54
Glu1855@OE1	Glu459@N	2.84	161.10	36.11
Glu1855@OE2	Arg434@NH1	2.79	154.89	31.93
Glu1205@OE2	Arg957@NH2	2.79	158.48	29.01
Glu1205@OE1	Arg957@NH2	2.79	158.76	25.54
Glu1205@OE1	Arg957@NH1	2.81	161.28	25.26
Glu1205@OE2	Arg957@NH1	2.81	160.99	23.90
Gln1797@O	Gln410@NE2	2.86	156.02	21.95
Glu1791@OE2	Asn914@ND2	2.83	163.02	21.05
Asp980@OD1	**Asn1206@ND2**	2.80	160.03	62.84
Leu485@O	**Ala2042@N**	2.86	158.97	56.60
Glu664@OE1	**Tyr1615@OH**	2.68	165.26	49.57
Glu617@OE1	**Ser1939@OG**	2.67	166.19	42.16
Asn712@O	**Arg1723@NH1**	2.80	155.32	35.57
Gln458@OE1	**Asn1856@N**	2.88	157.24	35.24
Asn712@O	**Arg1723@NE**	2.85	153.04	33.28
Glu459@OE1	**Ser1920@OG**	2.72	160.30	30.79
Glu410@OE1	**Gln1797@NE2**	2.85	162.12	28.28
Glu664@OE2	**Tyr1615@OH**	2.69	164.89	27.38
Glu459@OE2	**Ser1920@OG**	2.73	159.15	26.20
Glu617@OE1	**Ser1939@N**	2.87	159.01	24.82
Glu841@OE2	**Thr1783@OG1**	2.70	164.00	23.98
Glu459@OE2	**Val1857@N**	2.85	152.68	21.44

Material. With an acceptor-donor distance cut-off of <3.5 Å and angle cut off 120, the H-bond occupancy was calculated for the entire simulation trajectory. The H-bond analysis showed that an average of around 15-25 number of H-

bonds formed between the TLR4 and the vaccine construct (Figure 4.10.). It was also observed that the no. of hydrogen bonds increased after around 100 ns and was stable for the rest of the simulation. The pair of H-bond forming atoms with more than 20% occupancy and the corresponding parameters are listed in (**Table 4.9.)**.

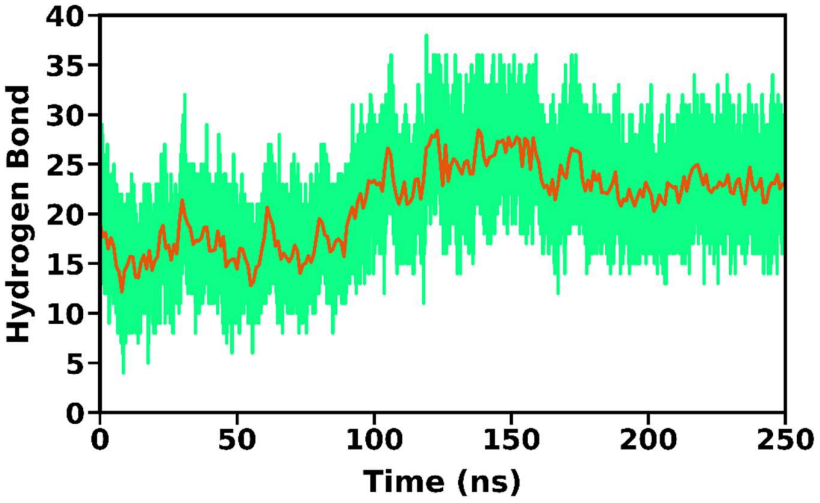
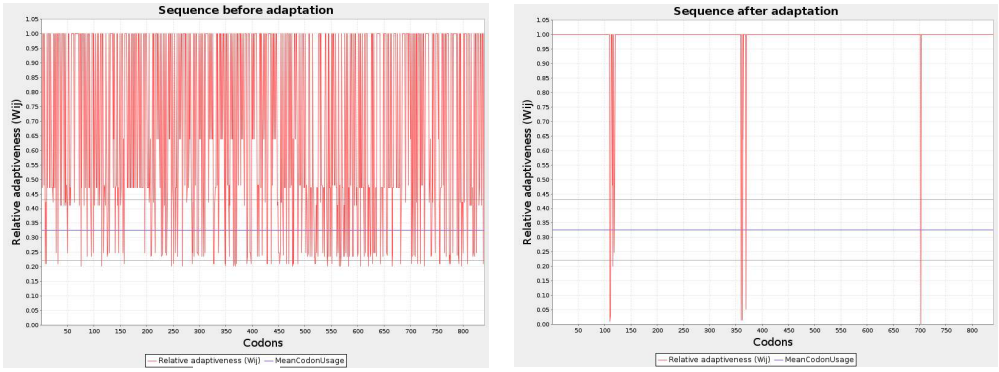


Figure 4.10. No. of Hydrogen bonds throughout the MD simulations.

4.3.14. Codon optimization and *In-silico* cloning:

In-silico cloning to work it was necessary to adapt the codon sequence of constructed vaccine to be expressed in the E.coli expression system. Therefore, for the incorporation of the vaccine construct in the plasmid vector, the 840 amino acid sequence of the vaccine construct is reverse translated into the cDNA of 2520 nucleotide. The resultant cDNA was codon optimized to ensure efficient expression of the constructed vaccine epitope in the Escherichia coli K12 host by using the JCAT server. To ensure efficient transcription to occur, exclusion of the rho-independent transcription termination site and restriction enzyme site from the cDNA sequence. Before adaptation, the cDNA sequence CAI value was 0.59 and its GC content was 56.4 %. After the codon adaptation, CAI value increases from 0.59 to 0.96 With 50.2 % of GC content. The increased CAI value shows that optimized cDNA has a majority of codons that are often used in Escherichia coli k12.



A B

Figure 4.11. Codon adaptation. *cDNA sequence (A) Before adaptation (B) After adaptation*

Next, cDNA of the constructed vaccine was reversed then at 5'site XhoI was added, and at 3'end BamHI was incorporated. By using SnapGene, the restriction enzyme BamHI and XhoI cleaved at their respective sites of the pET28a(+) expression vector. So, that cDNA can be inserted in between the cleaved site of the pET28a(+) vector, and for isolation and purification of vaccine construct 6 mer histidine tag was added to the C-terminal end. Vaccine gene depicted in blue color and present in between the sites of XhoI and BamHI (Figure 4.12).

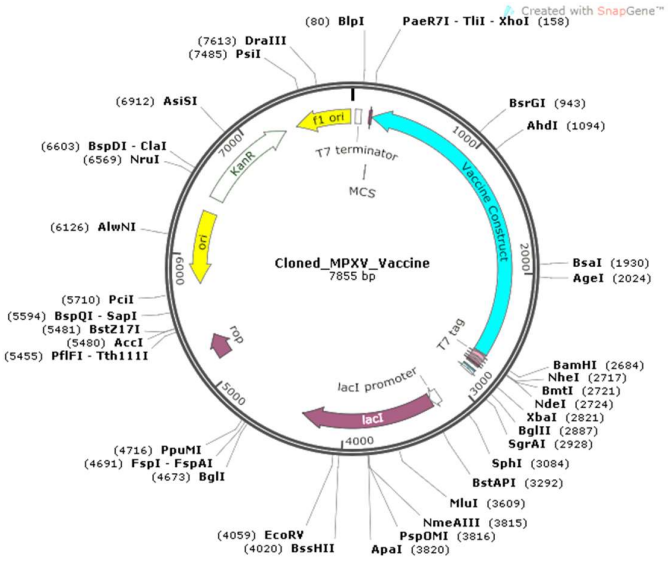


Figure 4.12. In silico cloning of construct vaccine in pET28a(+)

4.3.15. Computational Immune Assay:

A computer-generated simulation of the immune system was conducted to identify immunological responses to the developed immunization. The findings revealed that it was effective in eliciting immunological responses, as the increase in the production of humoral antibodies was responsible for the drop in antigen count. The rise in the peak of IgG1 + IgG2 indicates elicitation of primary response due to which antigen level decreased after each injection (>600,000 on the first dose and ~500,000 on the second and third doses), and it eventually leveled off on the fifth day after injection and this led to activation of secondary immunological response as the (Figure 4.13 A) data depicts the rise of IgM + IgG counts ~190,000 per ml in 10-15 days after each vaccination course. In response to injection, various cytokines such as IFN-γ, TGF-β, IL-10, and IFN-β levels have also risen shown in (Figure 4.13, A-B). These secondary immune responses led to tertiary responses in the body which cause the activation of a large number of B- cells population (Figure 4.14 A-C), similar trend has been shown with the activation of a large number of Helper T-cells (Figure 4.14 D-F) and cytotoxic T-cells (Figure A-B). These trends indicate that vaccination led to a successful secondary immunological response. It also led to an increase in the level of different types of innate immune responses (Figure 4.15 C-F). These results suggest that the design of multi-epitope vaccines has the capability to trigger a powerful immune response against Monkeypox Virus infection.

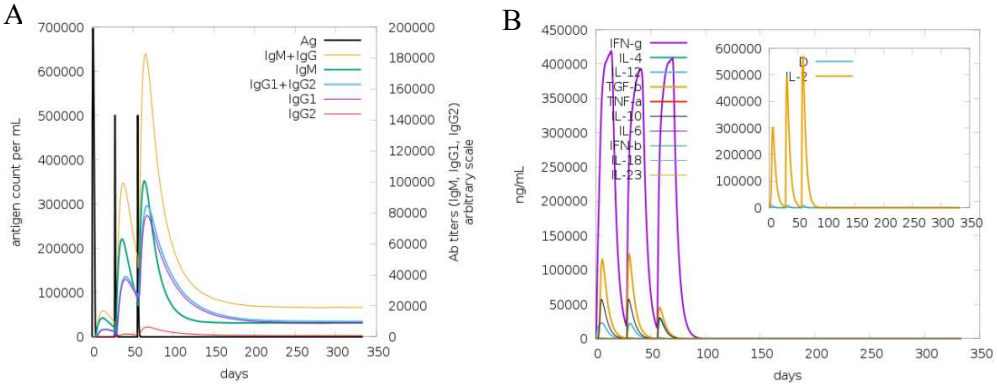


Figure 4.13. Antibody and cytokine responses to vaccination are predicted by the c-ImmSim service.

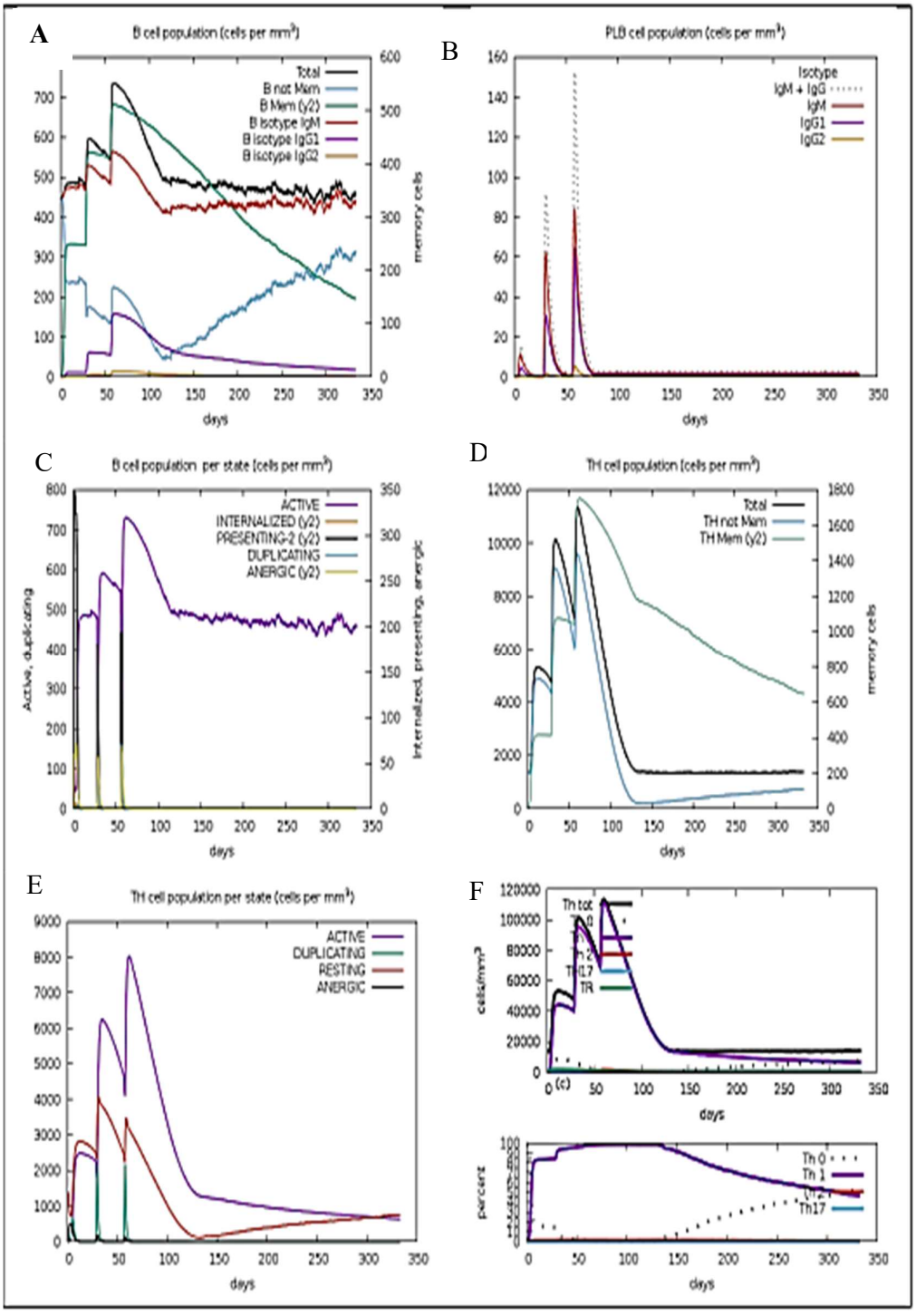


Figure 4.14. Plots (A-F) depicting responses of Immune simulation of B-cells and T- helper cells in response to the vaccine construct.

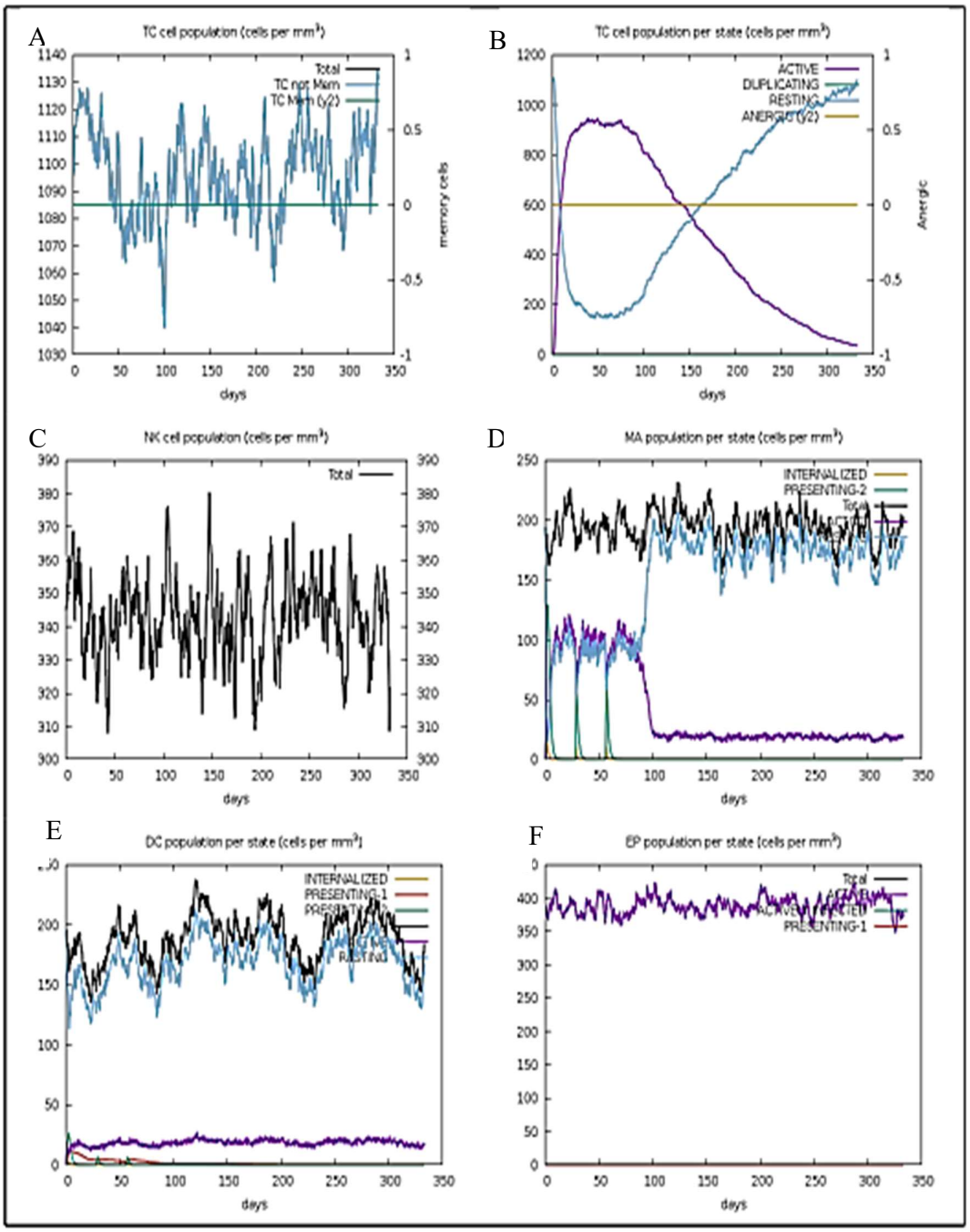


Figure 4.15. Plot showing (A and B) increase in cytotoxic T-cell population, and (C, D, E & F) an increase in the number of additional innate immune cells.

4.4 Conclusion

By using a variety of in-silico methods, a vaccine against Monkeypox virus was developed in the current work. To control this epidemic of viral infection, the availability of an effective vaccine is of utmost importance. There are various attempts made towards its vaccine development, but no successful vaccine with higher efficacy and efficiency is available to date. The advancement in the field of immunoinformatics analysis has overcome the disadvantage of traditional vaccine development. This study provides a brief immunoinformatics approach to find a novel vaccine candidate against MPXV virus infection. Here we explored the proteome of the MPXV virus to search for the best CTL, HTL, and B cell epitopes to construct a multiepitope subunit vaccine. Sequence analysis of the vaccine construct showed the predicted epitopes are antigenic and non-allergic in nature. MD simulation depicted the stability of the tertiary structure of the vaccine construct and its complex with TLR4 complex and in silico cloning analysis helped in the generation of an effective vector that can be used for the multi-epitope vaccine expression. In summary, an extensive immunoinformatics approach was exploited to develop an effective and efficacious multi-epitope vaccine against the Monkeypox virus.

Chapter 5

Conclusion and Future Perspective

5.1 Conclusion of the Thesis

We investigated therapeutic options for bacterial diseases in this study. Bacterial diseases have a significant impact on both global health and the economy. The current thesis work is focused on exploring and targeting the novel G-quadruplex motif as a potential drug target for antibacterial therapy. The prevalence of GQs in viruses has only recently been investigated, but the results are quite fascinating. We addressed this aspect in the Monkeypox virus with twelve distinct genome types. We identified three highly conserved G-quadruplex forming motifs (HPGQs) and seven have a prevalence amount of G-quadruplex motif and are characterized by various biophysical assays, including NMR, CD spectroscopy, and EMSA assay, all of which showed the formation of stable GQs. Binding analysis using ITC and melting analysis using CD revealed a high affinity of a GQ-specific ligand, this result further confirmed that the GQ ligand interacts with the GQ structure in the monkeypox virus genome and may have a regulatory role in gene regulation.

The last objective of the thesis was to develop an effective and efficacious multi-epitope vaccine against the Monkeypox virus. There are various attempts made towards its vaccine development, but no successful vaccine with higher efficacy and efficiency is available to date. The development of vaccines can help to control the spread of the virus and prevent new cases from occurring. By using a variety of in-silico methods, a vaccine against Monkeypox virus was developed in the current work.

In summary, the thesis work focused on using Monkeypox virus genome to identify and characterizes the G-quadruplexes motif to design anti-viral drug therapies and target. The GQ-forming sequences predicted in Monkeypox genomes were found to interact with GQ-specific ligands and may have potential modulatory roles in gene expression. Our comprehensive in silico

research offers a foundation for the creation of a Monkeypox virus vaccine as a prevention measure.

.

5.2 Future aspects

GQs have recently been identified as potential therapeutic targets against a variety of human pathogens, including viruses, bacteria, and protozoa. The revelation of GQ sequences linked to various diseases should go hand in hand with the development of GQ-targeting ligands. GQ motifs have small molecule binding sites, and their interaction can lead to the stabilization of complexes that regulate several biological processes. In this study, Cell-based assay studies are required to further validate the biological significance of the predicted GQ motif in the regulation of the expression of associated genes. That aids the researcher's confidence that GQ-targeted small molecules have a promising future in the treatment of infectious diseases. Our comprehensive in silico development of vaccines offers a foundation for the creation of a Monkeypox virus vaccine as a prevention measure.

Appendix 1

Appendix 1A. Per-residue decomposition of binding free energy (≤ -1 kcal/mol)

Vaccine	**Energy (kcal/mol)**	**TLR4**	**Energy (kcal/mol)**
Thr1783	-4.58	Arg434	-6.98
Ser1939	-4.52	Gln458	-5.47
Ile1721	-4.49	Asn712	-5.13
Lys1820	-4.30	Lys534	-4.84
Met1781	-4.23	Phe983	-3.04
Leu1227	-3.78	Val605	-2.78
Tyr1720	-3.77	Pro688	-2.70
Arg1723	-3.60	Arg957	-2.65
Lys1935	-3.47	Pro487	-2.44
Pro1804	-3.36	Hid1104	-2.40
Tyr1615	-3.30	Leu618	-2.33
Lys1854	-3.24	Gln666	-2.19
Tyr1782	-3.10	Phe461	-2.10
Asn1206	-3.06	Glu413	-2.08
Val1857	-2.92	Leu736	-1.93
Trp1859	-2.88	Glu841	-1.83
Gln1797	-2.81	Ser956	-1.78
Leu1717	-2.78	Gln410	-1.72
Tyr1802	-2.73	Asp784	-1.67
Met1203	-2.65	Asn840	-1.67
Lys1924	-2.62	Asp756	-1.61
Met1806	-2.61	Glu837	-1.57
Tyr1787	-2.42	Glu911	-1.48
Val1230	-2.39	Asp869	-1.44
Lys1917	-2.37	Glu717	-1.38
Lys1866	-2.36	Asp369	-1.38
Lys1918	-2.36	Glu896	-1.37
Lys1818	-2.29	Glu845	-1.33
Asn2041	-2.29	Asn914	-1.32
Ile1722	-2.18	Glu729	-1.32
Lys1936	-2.09	Glu606	-1.31
Arg2037	-1.98	Asp813	-1.30
Lys1864	-1.96	Glu468	-1.28
Ile1234	-1.96	Asp769	-1.28
Lys1213	-1.95	Glu669	-1.24
Lys1989	-1.94	Glu343	-1.23
Arg1796	-1.94	Glu862	-1.20
Arg1312	-1.93	Asp970	-1.19

Lys1972	-1.84	Glu396	-1.19
Lys1971	-1.78	Glu944	-1.17
Lys1610	-1.73	Leu787	-1.13
Arg1885	-1.71	Gln1080	-1.13
Arg2032	-1.62	Glu753	-1.11
Ala1801	-1.62	Glu1000	-1.11
Arg1251	-1.60	Asp873	-1.10
Lys1882	-1.60	Asp345	-1.10
Lys1874	-1.59	Asp900	-1.10
Lys1908	-1.58	Glu1049	-1.08
Arg1967	-1.58	Asp848	-1.08
Asn1856	-1.58	Glu744	-1.07
Lys1990	-1.56	Asp946	-1.07
Lys1998	-1.56	Glu951	-1.06
Lys1899	-1.53	Asp670	-1.06
Lys2026	-1.49	Asp874	-1.05
Lys2003	-1.47	Glu537	-1.01
Lys1568	-1.44	Glu997	-1.01
Arg1959	-1.44		
Leu1128	-1.42		
Lys1900	-1.41		
Arg1238	-1.41		
Lys2025	-1.40		
Arg2016	-1.40		
Lys1619	-1.40		
Lys1611	-1.39		
Lys1609	-1.38		
Lys2013	-1.38		
Arg1240	-1.36		
Arg1560	-1.35		
Arg1776	-1.34		
Arg1303	-1.34		
Lys1954	-1.34		
Leu1798	-1.33		
Arg1296	-1.32		
Lys1953	-1.32		
Arg1245	-1.32		
Arg1323	-1.31		
Lys1364	-1.31		
Lys1508	-1.31		
Arg1258	-1.30		
Ile2040	-1.27		
Lys2008	-1.26		
Arg2015	-1.25		
Arg1310	-1.25		
Lys1577	-1.25		
Lys1772	-1.24		
Glu1178	-1.24		

Lys1574	-1.23		
Lys1620	-1.23		
Arg1249	-1.22		
Glu1180	-1.21		
Arg2018	-1.20		
Lys1730	-1.19		
Lys1581	-1.17		
Lys1363	-1.16		
Lys2007	-1.16		
Lys1666	-1.15		
Arg1256	-1.12		
Lys1261	-1.11		
Ala1204	-1.11		
Lys1575	-1.10		
Arg1274	-1.10		
Gln1795	-1.09		
Lys1516	-1.09		
Arg1441	-1.08		
Lys1515	-1.08		
Arg1764	-1.06		
Thr1833	-1.05		
Lys1578	-1.05		
Lys1284	-1.05		
Lys1355	-1.04		
Lys1375	-1.03		
Lys1369	-1.03		
Lys1370	-1.02		
Arg1347	-1.01		

References:

[1] Huang, Y., Mu, L., & Wang, W. (2022). Monkeypox: epidemiology, pathogenesis, treatment and prevention. Signal transduction and targeted therapy, 7(1), 373. https://doi.org/10.1038/s41392-022-01215-4

[2] Kaler, J., Hussain, A., Flores, G., Kheiri, S., & Desrosiers, D. (2022). Monkeypox: A Comprehensive Review of Transmission, Pathogenesis, and Manifestation. Cureus, 14(7), e26531. https://doi.org/10.7759/cureus.26531

[3] 3.Wenham, C., & Eccleston-Turner, M. (2022). Monkeypox as a PHEIC: implications for global health governance. *Lancet (London, England)*, *400*(10369), 2169–2171. https://doi.org/10.1016/S0140-6736(22)01437-4

[4] Nolen, L. D., Osadebe, L., Katomba, J., Likofata, J., Mukadi, D., Monroe, B., Doty, J., Hughes, C. M., Kabamba, J., Malekani, J., Bomponda, P. L., Lokota, J. I., Balilo, M. P., Likafi, T., Lushima, R. S., Ilunga, B. K., Nkawa, F., Pukuta, E., Karhemere, S., Tamfum,J. J., … Reynolds, M. G. (2016). Extended Human-to-Human Transmission during aMonkeypox Outbreak in the Democratic Republic of the Congo. Emerging infectious diseases, 22(6), 1014–1021. https://doi.org/10.3201/eid2206.150579

[5] Zhu, M., Ji, J., Shi, D., Lu, X., Wang, B., Wu, N., Wu, J., Yao, H., & Li, L. (2022). Unusual global outbreak of monkeypox: what should we do?. *Frontiers of medicine*, *16*(4), 507–517. https://doi.org/10.1007/s11684-022-0952-z

[6] Hazel, P., Huppert, J., Balasubramanian, S., & Neidle, S. (2004). Loop-length-dependent folding of G-quadruplexes. *Journal of the American Chemical Society*, *126*(50), 16405–16415. https://doi.org/10.1021/ja045154j

[7] Burge, S., Parkinson, G. N., Hazel, P., Todd, A. K., & Neidle, S. (2006). Quadruplex DNA: sequence, topology and structure. Nucleic acids research, 34(19), 5402–5415. https://doi.org/10.1093/nar/gkl655

[8] Krupovic, M., Dolja, V. V., & Koonin, E. V. (2019). Origin of viruses: primordial replicators recruiting capsids from hosts. Nature reviews. Microbiology, 17(7), 449–458. https://doi.org/10.1038/s41579-019-0205-6

[9] Perrone, R., Nadai, M., Poe, J. A., Frasson, I., Palumbo, M., Palù, G., Smithgall, T. E., & Richter, S. N. (2013). Formation of a unique cluster of G-quadruplex structures in the HIV-1 Nef coding region: implications for antiviral activity. PloS one, 8(8), e73121. https://doi.org/10.1371/journal.pone.0073121

[10] Norseen, J., Johnson, F. B., & Lieberman, P. M. (2009). Role for G-quadruplex RNA binding by Epstein-Barr virus nuclear antigen 1 in DNA replication and metaphase chromosome attachment. Journal of virology, 83(20), 10336–10346. https://doi.org/10.1128/JVI.00747-0925

[11] Biswas, B., Kandpal, M., & Vivekanandan, P. (2017). A G-quadruplex motif in an envelope gene promoter regulates transcription and virion secretion in HBV genotype B.Nucleic acids research, 45(19), 11268–11280. https://doi.org/10.1093/nar/gkx823

[12] Majee, P., Pattnaik, A., Sahoo, B. R., Shankar, U., Pattnaik, A. K., Kumar, A., & Nayak, D. (2021). Inhibition of Zika virus replication by G-quadruplex-binding ligands. Molecular therapy. Nucleic acids, 23, 691–701. https://doi.org/10.1016/j.omtn.2020.12.030

[13] Wang, S. R., Zhang, Q. Y., Wang, J. Q., Ge, X. Y., Song, Y. Y., Wang, Y. F., Li, X. D., Fu, B. S., Xu, G. H., Shu, B., Gong, P., Zhang, B., Tian, T., & Zhou, X. (2016). Chemical Targeting of a G-Quadruplex RNA in the Ebola Virus L Gene. Cell chemical biology, 23(9), 1113–1122. https://doi.org/10.1016/j.chembiol.2016.07.019

[14] Majee, P., Kumar Mishra, S., Pandya, N., Shankar, U., Pasadi, S., Muniyappa, K., Nayak, D., & Kumar, A. (2020). Identification and characterization of two conserved G- quadruplex forming motifs in the Nipah virus genome and their interaction with G-quadruplex specific ligands. Scientific reports, 10(1), 1477. https://doi.org/10.1038/s41598-020-58406-8

[15] Lim, S. Y., Osuna, C. E., Best, K., Taylor, R., Chen, E., Yoon, G., Kublin, J. L., Schalk, D., Schultz-Darken, N., Capuano, S., Safronetz, D., Luo, M., MacLennan, S., Mathis, A.,Babu, Y. S., Sheridan, W. P., Perelson, A. S., & Whitney, J. B. (2020). A direct-acting antiviral drug abrogates viremia in Zika virus-infected rhesus macaques. Science translational medicine,12(547), eaau9135. https://doi.org/10.1126/scitranslmed.aau9135

[16] Bhartiya, D., Chawla, V., Ghosh, S., Shankar, R., & Kumar, N. (2016). Genome-wide regulatory dynamics of G-quadruplexes in

human malaria parasite Plasmodium falciparum. Genomics, 108(5-6), 224–231. https://doi.org/10.1016/j.ygeno.2016.10.004

[17] Cahoon, L. A., & Seifert, H. S. (2009). An alternative DNA structure is necessary for pilin antigenic variation in Neisseria gonorrhoeae. Science (New York, N.Y.), 325(5941), 764–767. https://doi.org/10.1126/science.1175653

[18] Vorlíčková, M., Kejnovská, I., Sagi, J., Renčiuk, D., Bednářová, K., Motlová, J., & Kypr, J. (2012). Circular dichroism and guanine quadruplexes. Methods (San Diego, Calif.), 57(1), 64–75. https://doi.org/10.1016/j.ymeth.2012.03.011

[19] Adrian, M., Heddi, B., & Phan, A. T. (2012). NMR spectroscopy of G-quadruplexes. Methods (San Diego, Calif.), 57(1), 11–24. https://doi.org/10.1016/j.ymeth.2012.05.003

[20] Ward, W. H., & Holdgate, G. A. (2001). Isothermal titration calorimetry in drug discovery. Progress in medicinal chemistry, 38, 309–376. https://doi.org/10.1016/s0079-6468(08)70097-326

[21] Holden, N. S., & Tacon, C. E. (2011). Principles and problems of the electrophoretic mobility shift assay. Journal of pharmacological and toxicological methods, 63(1), 7–14. https://doi.org/10.1016/j.vascn.2010.03.002

[22] Kollman, P. A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., Lee, M., Lee, T.,Duan, Y., Wang, W., Donini, O., Cieplak, P., Srinivasan, J., Case, D. A., & Cheatham, T.E., 3 rd (2000). Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. Accounts of chemical research, 33(12), 889–897. https://doi.org/10.1021/ar000033j

[23] Shankar, U., Jain, N., Mishra, S. K., Sk, M. F., Kar, P., & Kumar, A. (2022). Mining of Ebola virus genome for the construction of multi-epitope vaccine to combat its infection. Journal of biomolecular structure & dynamics, 40(11), 4815–4831. https://doi.org/10.1080/07391102.2021.1874529

[24] Roy, R., Mishra, A., Poddar, S., Nayak, D., & Kar, P. (2022). Investigating the mechanism of recognition and structural dynamics of nucleoprotein-RNA complex from Peste des petits ruminants virus via Gaussian accelerated molecular dynamics simulations. Journal of biomolecular structure & dynamics, 40(5), 2302–2315. https://doi.org/10.1080/07391102.2020.1838327

[25] Roy, R., Ghosh, B., & Kar, P. (2020). Investigating Conformational Dynamics of Lewis Y Oligosaccharides and Elucidating Blood Group Dependency of Cholera Using Molecular Dynamics. ACS omega, 5(8), 3932–3942. https://doi.org/10.1021/acsomega.9b03398

[26] Gohlke, H., Kiel, C., & Case, D. A. (2003). Insights into protein-protein binding by binding free energy calculation and free energy decomposition for the Ras-Raf and Ras-RalGDS complexes. Journal of molecular biology, 330(4), 891–913. https://doi.org/10.1016/s0022-2836(03)006107

[27] Lv, L., & Zhang, L. (2022). Characterization of G-Quadruplexes in Enterovirus A71 Genome and Their Interaction with G-Quadruplex Ligands. Microbiology spectrum,10(3), e0046022. https://doi.org/10.1128/spectrum.00460-22

[28] Majee, P., Shankar, U., Pasadi, S., Muniyappa, K., Nayak, D., & Kumar, A. (2020).Genome-wide analysis reveals a regulatory role for G-quadruplexes during Adenovirus multiplication. Virus research, 283, 197960. https://doi.org/10.1016/j.virusres.2020.197960

[29] Tomaszewska, M., Szabat, M., Zielińska, K., & Kierzek, R. (2021). Identification and Structural Aspects of G-Quadruplex-Forming Sequences from the Influenza A Virus Genome. International journal of molecular sciences, 22(11), 6031. https://doi.org/10.3390/ijms2211603127

[30] Fleming, A. M., Ding, Y., Alenko, A., & Burrows, C. J. (2016). Zika Virus Genomic RNA Possesses Conserved G-Quadruplexes Characteristic of the Flaviviridae Family. ACS infectious diseases, 2(10), 674–681. https://doi.org/10.1021/acsinfecdis.6b00109

[31] Zahin, M., Dean, W. L., Ghim, S. J., Joh, J., Gray, R. D., Khanal, S., Bossart, G. D., Mignucci-Giannoni, A. A., Rouchka, E. C., Jenson, A. B., Trent, J. O., Chaires, J. B., & Chariker, J. H. (2018). Identification of G-quadruplex forming sequences in three manatee papillomaviruses. PloS one, 13(4), e0195625. https://doi.org/10.1371/journal.pone.0195625

[32] Deng, H., Gong, B., Yang, Z., Li, Z., Zhou, H., Zhang, Y., Niu, X., Liu, S., & Wei, D. (2019). Intensive Distribution of G_2-Quaduplexes in the Pseudorabies Virus Genome and Their Sensitivity to Cations and G-Quadruplex Ligands. Molecules (Basel, Switzerland), 24(4), 774. https://doi.org/10.3390/molecules24040774

[33] Vorlíčková, M., Kejnovská, I., Sagi, J., Renčiuk, D., Bednářová, K., Motlová, J., & Kypr, J. (2012). Circular dichroism and guanine quadruplexes. Methods (San Diego, Calif.),57(1), 64–75. https://doi.org/10.1016/j.ymeth.2012.03.011

[34] Wang, Z., & Liu, J. P. (2017). Effects of the central potassium ions on the G-quadruplexand stabilizer binding. Journal of molecular graphics & modelling, 72, 168–177. https://doi.org/10.1016/j.jmgm.2017.01.006

[35] Petersen, E., Kantele, A., Koopmans, M., Asogun, D., Yinka-Ogunleye, A., Ihekweazu, C., & Zumla, A. (2019). Human Monkeypox: Epidemiologic and Clinical Characteristics, Diagnosis, and Prevention. *Infectious disease clinics of North America*, *33*(4), 1027–1043. https://doi.org/10.1016/j.idc.2019.03.001

[36] 37. Babkin, I. V., Babkina, I. N., & Tikunova, N. V. (2022). An Update of Orthopoxvirus Molecular Evolution. *Viruses*, *14*(2), 388. https://doi.org/10.3390/v14020388

[37] 38. Shchelkunova, G. A., & Shchelkunov, S. N. (2022). Smallpox, Monkeypox and Other Human Orthopoxvirus Infections. *Viruses*, *15*(1), 103. https://doi.org/10.3390/v15010103

[38] 39. Pandey, R. K., Bhatt, T. K., & Prajapati, V. K. (2018). Novel Immunoinformatics Approaches to Design Multi-epitope Subunit Vaccine for Malaria by Investigating Anopheles Salivary Protein. *Scientific reports*, *8*(1), 1125. https://doi.org/10.1038/s41598-018-19456-1

[39] Pandya, N., & Kumar, A. (2023). Immunoinformatics analysis for design of multi-epitope subunit vaccine by using heat shock proteins against *Schistosoma mansoni*. *Journal of biomolecular structure & dynamics*, *41*(5), 1859–1878. https://doi.org/10.1080/07391102.2021.2025430

[40] Doytchinova, I. A., & Flower, D. R. (2007). VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC bioinformatics*, *8*, 4. https://doi.org/10.1186/1471-2105-8-4

[41] El-Manzalawy, Y., Dobbs, D., & Honavar, V. (2008). Predicting linear B-cell epitopes using string kernels. *Journal of molecular recognition : JMR*, *21*(4), 243–255. https://doi.org/10.1002/jmr.893

[42] Nezafat, N., Ghasemi, Y., Javadi, G., Khoshnoud, M. J., & Omidinia, E. (2014). A novel multi-epitope peptide vaccine against cancer: an in silico approach. *Journal of theoretical biology*, *349*, 121–134. https://doi.org/10.1016/j.jtbi.2014.01.018

[43] Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D., & Bairoch, A. (2003). ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic acids research*, *31*(13), 3784–3788. https://doi.org/10.1093/nar/gkg563

[44] Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of applied crystallography*, *26*(2), 283-291.

[45] Rapin, N., Lund, O., Bernaschi, M., & Castiglione, F. (2010). Computational immunology meets bioinformatics: the use of prediction tools for molecular binding in the simulation of the immune system. *PloS one*, *5*(4), e9862. https://doi.org/10.1371/journal.pone.0009862

[46] Sami, S. A., Marma, K. K. S., Mahmud, S., Khan, M. A. N., Albogami, S., El-Shehawi, A. M., Rakib, A., Chakraborty, A., Mohiuddin, M., Dhama, K., Uddin, M. M. N., Hossain, M. K., Tallei, T. E., & Emran, T. B. (2021). Designing of a Multi-epitope Vaccine against the Structural Proteins of Marburg Virus Exploiting the Immunoinformatics Approach. *ACS omega*, *6*(47), 32043–32071. https://doi.org/10.1021/acsomega.1c04817

[47] Jung, I. D., Jeong, S. K., Lee, C. M., Noh, K. T., Heo, D. R., Shin, Y. K., Yun, C. H., Koh, W. J., Akira, S., Whang, J., Kim, H. J., Park, W. S., Shin, S. J., & Park, Y. M. (2011). Enhanced efficacy of therapeutic cancer vaccines produced by co-treatment with Mycobacterium tuberculosis heparin-binding hemagglutinin, a novel TLR4 agonist. *Cancer research*, *71*(8), 2858–2870. https://doi.org/10.1158/0008-5472.CAN-10-3487