Analysis of Codon Usage Bias in Canine Circovirus and In Silico Approach for Prediction of Vaccine Candidate

Thesis Submitted for the Award of the Degree of

DOCTOR OF PHILOSOPHY

in

BIOTECHNOLOGY

By

PANKAJ JAIN

(41900145)

Supervised By

Dr. Vikas Kaushik

LOVELY PROFESSIONAL UNIVERSITY
PUNJAB
Nov 2022

DECLARATION

I hereby declare that the thesis entitled, "***Analysis of Codon Usage Bias in Canine Circovirus and In Silico Approach for Prediction of Vaccine Candidate***" was submitted for Ph.D. Biotechnology degree to Department of Biotechnology, Lovely Professional University is entirely original work and all ideas and references have been duly acknowledged. The research work has not formed the basis for the award of any other degree.

Date: 21st Nov, 2022

Pankaj Jain

Registration No: **41900145**

CERTIFICATE

This is to certify that **Mr. Pankaj Jain** has completed the Ph.D. in Biotechnology titled "***Analysis of Codon Usage Bias in Canine Circovirus and In Silico Approach for Prediction of Vaccine Candidate***" under my guidance and supervision. To the best of my knowledge, the present work is the result of his original investigation and study. No part of this thesis has ever been submitted for any other degree or diploma.

The thesis is fit for the submission for the partial fulfillment of the condition for the award of a degree of Ph.D. in Biotechnology.

Date: 21st Nov, 2022 **Advisor Name & Signature: Vikas Kaushik**

Acknowledgments

I would like to thank my esteemed supervisor, Dr. Vikas Kaushik, for his invaluable supervision, support, and tutelage during my Ph.D. degree. His guidance and advice carried me through all the stages of writing my thesis. I would also like to thank my committee members for letting my defense be an enjoyable moment, and for your brilliant comments and suggestions, thanks to you.

I would also like to thank Amit Joshi, Sunil Krishnan, Nahid Akhtar, Dr. Umesh, Dr. Neeta Raj Sharma, Dr. Rekha, Dr. Nitin Tandon, and the faculty members of the School of Bioengineering and Biotechnology for their valuable guidance throughout my studies. You provided me with the tools that I needed to choose the right direction and complete my dissertation.

Special thanks to my mother, father, my loving wife Shikha, my daughters Sanya and Vanya, and my family as a whole for their continuous support and understanding when undertaking my research and writing my thesis. Your prayer for me was what sustained me this far.

Finally, I would like to thank God, for letting me through all the difficulties. I have experienced your guidance day by day. You are the one who let me finish my degree. I will keep on trusting you for my future.

TABLE OF CONTENTS

Sr No.	Chapter Title	Page No.
1.	**Chapter 1: Introduction**	1- 4
2.	**Chapter 2: Review and Literature**	5-28
3.	**Chapter 3: Objectives and Scope of Study**	29-31
4.	**Chapter 4: Methods and Materials**	32-47
	1.1. Analysis of synonymous codon usage characteristics of CanineCV	33-34
	1.1.1 The complete genomes of CanineCV strains	
	1.1.2 Nucleotide Composition Analysis	
	1.1.3 Relative Synonymous Codon Usage	
	1.1.4 Effective number of codons (ENC)	
	1.1.5 Codon dinucleotide frequency analysis	
	1.1.6 Gravy and aroma statistics	
	2.1. Evaluate factors such as mutational pressure and natural selection responsible for the host-specific adaptation of CanineCV	34-37
	2.1.1 Enc-plot analysis	
	2.1.2 Parity rule 2 (PR2) analysis	
	2.1.3 Neutrality analysis	
	2.1.4 Analysis of host-specific adaptation	
	2.1.5 Correspondence Analysis (CoA)	
	2.1.6 Statistical Analysis	
	3.1. In-silico prediction of multi-epitope vaccine construct	37-39
	3.1.1 Proteomic data retrieval	
	3.1.2 Epitopes selection and structure prediction	
	3.1.3 Molecular docking and simulation at epitope-DLA level	
	3.1.4 Full-fledged in-silico vaccine construction	
	3.1.5 Vaccine construct prediction and affirmation	
	3.1.6 Codon Adaptation and vaccine in-silico cloning	
	3.2. In-silico Epitope based vaccine prediction and wet-lab validations	39-46
	3.2.1 Protein Sequence Retrieval	
	3.2.2 Antigenicity, Toxicity, Conservancy, and IFN-γ Analysis	
	3.2.3 Vaccine Engineering and Determination of its Physiochemical Properties	
	3.2.4 Molecular Modeling, Docking, and Molecular Dynamic Simulations of Designed Multi Epitope Vaccine with TLR-5	
	3.2.5 Peptide Synthesis and Conjugation with Carrier Protein	
	3.2.6 In-vivo Validation for Antibody Development and Assessment	
	4.1. Codon usage bias manipulation strategy for the development of a vaccine	46-47
	4.1.1 Four strategies are used to generate genome-scale deoptimized viruses	

5. **Chapter 5: Results and Discussion** 48-87
5.1. In-silico Prediction of multi-epitope vaccine construct 49-66
5.1.1 Proteome structure and sequence 55-56
5.1.2 Predicted epitopes and DLA structure 56-57
5.1.3 Molecular Docking and MD simulation 57-59
5.1.4 Full-fledged vaccine construction 59-60
5.1.5 Predicted secondary and tertiary structure 60-62
5.1.6 Codon Adaptation and vaccine in-silico cloning 62-66
5.2. In-silico Epitope based vaccine prediction and wet-lab validations 67-87
5.2.1 Prediction of Antigenic Epitopes and Analysis of their Antigenicity 67-69
5.2.2 Canine CV Vaccine Formulation and Determination of its Physiochemical Properties 69-70
5.2.3 Modeling and Docking of TLR-5 and Multi-epitope Vaccine Construct 71-76
5.2.4 Structural Stability of TLR-5 and MEV complex 77
5.2.5 Detailed Analysis of Snapshots obtained during MD Simulations and its Comparison with Initial Modeled TLR5-MEVConstruct 81
5.2.6 Peptide Synthesis and Conjugation to Carrier Protein
5.2.7 In-vivo Validation for Antibody Development and Assessment 81-87
6. **Chapter 6: Summary and Conclusion** 88-92
7. **Bibliography** 93-107
8. **Supplementary Table** *S1

LIST OF TABLES

Table No.	Table Title	Page No.
1.	Methodology Used	37
2.	Canine Circovirus variants sequences used for this study	49
3.	Nucleotides composition of CanineCV genome at third codon position	*S1
4.	RSCU values	53
5.	Retrieved FASTA sequences	56
6.	Epitope selection based on Epitope prediction score and antigenicity score	56-57
7.	DLA-Epitope complex docking analysis	59
8.	The full-fledged vaccine sequence	59
9.	Physiochemical Properties of the full-fledged vaccine	59-60
10.	Codon adaptation index for the gene of a finalized full-fledged vaccine against	63
11.	Epitope screening based on VaxiJen score	68
12.	List of the antigenic epitopes based on the GenScript optimum antigen tool	68-69

LIST OF FIGURES

Fig No.	Figure Legend	Page No.
1.	Nc-GC3 plot of CanineCV	55
2.	Docked complexes	58
3.	RMSD plot for 5 selected epitopes-DLA complexes	59
4.	Full-fledged vaccine sequence and secondary structure description	61
5.	The 3D structure of vaccine construct	61
6.	Ramachandran plot for a full-fledged vaccine against Canine circovirus	62
7.	Vaccine constructs in pET28a (+) vector	64
8.	Schematic representation of the predicted multi-epitope vaccine (MEV)	70
9.	The modeled 3D structures of the a) MEV construct and b) Cartoon representation of a modeled Toll-like Receptor 5 (TLR-5)	73
10.	(Upper Panel) The docked complex of MEV construct (green cartoon) and TLR-5 (marine blue cartoon)	74
11.	a. RMSD of TLR-5 Time evolution of backbone and MEV construct during MD simulations b) Backbone RMSF plots c) Time evolution of the number of hydrogen bonds between TLR-5 and MEV construct d) Variation in Interaction energy plot with segregated electrostatics (E-Coul) and van der Waals (E-LJ) component e) Solvent accessible surface area of TLR-5 and MEV construct	75
12.	a) Superimposition of selected snapshots of TLR-5 and MEV construct b) Contact maps showing the conservation of contacts between residues in between TLR-5 and Vaccine construct	76
13.	MD cons consensus map of 1000 MD snapshots	80
14	ELISA assay scatter plot of OD_{450} values and serum dilutions from (1:1000 to 1:64,000) of CanineCV antibody and IgG control with varying concentrations of antigen. The OD_{450} values can be inversely correlated to serum dilutions.	82

LIST OF ABBREVIATIONS

NCBI	National Centre for Biotechnology Information
EMBL	European Molecular Biology Laboratory
BLAST	Basic Local Alignment Search Tool
RCSB	Research Collaboratory for Structural Bioinformatics
EBV	Epitope Based Vaccine
MEV	Multiple Epitope Vaccine
RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuation
CCV	Canine Circovirus
REP	Replicase
CAP	Capsid
CV	Circovirus
IFN	Interferon
IL	Interleukin
MHC	Major Histocompatibility Complex
HLA	Human Leukocyte Antigen
ANN	Artificial Neural Network
HMM	Hidden Markov Models
MD	Molecular Dynamics
RSCU	Relative Synonymous Codon Usage
CUB	Codon Usage Bias
PDB	Protein Databank
ENC	Effective No. of Codons
GC	Guanine Cytosine content
CAI	Codon Adaptive Index
DLA	Dog Leukocyte Antigen
ORF	Open Reading Frame
TLR	Toll Like Receptor

Abstract

Canine circovirus is a life-threatening disease that causes vasculitis and haemorrhagic enteritis in dogs. In pigs, foxes, and dogs, it causes fatal gastroenteritis. In the dog circovirus genome, two major (and opposing) transcriptional units contain 2 genes encoding (Exons), a replicase protein (Rep), as well as the capsid (Cap) proteins. 303 amino acids make up the replicase protein and 270 amino acids make up the capsid protein. Nucleotide composition and codon usage bias of *cap* & *rep* gene of 88 CaCV strains were compared. The overall nucleotide composition of CDSs at the third codon position (A3% C3% G3% T3%) along with overall GC% and AT%, GC12, and GC3 were analyzed. The Aroma and Gravy values, effective no of codons (ENC), and relative synonymous codon usage (RSCU) values were also analyzed using codon W 1.4.2 and CAIcal server. The results show that CaCV genome is AT rich and A/T ending codons were preferred then GC-ending codons. Nc-GC3 plot of CanineCV reveals that selection pressure was dominant over mutational pressure. The correlation analysis between CAI, Gravy, and Aroma indicates natural selection over mutational pressure. The RSCU values of *Cap* & *Rep* genes were analyzed to find out overrepresented codons. Scientists used immunoinformatics techniques including epitope scanning, molecular modeling, and MD simulations to develop a peptide vaccine against Canine circovirus. The vaccine was created by using an appropriate linker to connect the targeted epitopes with adjuvants. In silico technologies were also used for cloning and expression of the vaccine construct. The NetMHC server, which employs the ANN (Artificial Neural Networking) algorithm, was used to screen epitopes. Such methods for identifying epitopes that bind particular dog leukocyte antigens (DLA) and thus elicit a protective immunity are simple and low-cost. Overall, five epitopes were chosen and utilized to develop a vaccine construct: YQHLPPFRF, YIRAKWINW, ALYRRLTLI, HLQGFVNLK, and GTMNFVARR. According to molecular docking and molecular dynamics simulations, these epitopes can attach to DLA molecules with stability. Based on In-Silico engineering and codon adaption tests, the vaccine may be generated by the E. coliK12 variant. The data suggest that the vaccination design might help to reduce Canine circovirus outbreaks in dogs. To corroborate the results, additional in vitro and in vivo research was required. As a result, we conducted fresh research with additional validations. In this work, we developed a potential multi-epitope

vaccination (MEV) construct that targets several CanineCV strains. Capsid and replicase proteins from four distinct strains of CanineCV were used to predict MHCII binding CD4+T cell epitope peptides. Among them, five conserved epitope peptides among the four CanineCV strains were selected. The final vaccine was constructed using antigenic, non-toxic, and conserved multiple epitopes identified *in-silico*. Further, the molecular docking and molecular dynamics predicted stable interactions between predicted MEV and canine receptor TLR-5. To validate antigenicity and immunogenicity, one of the mapped epitope peptides was synthesized. The *in-vivo* analysis of the selected epitope indicates $CD4^{+}$T cell-dependant generation of antibodies which further suggests that the MEV construct designed holds the promise of a candidate for a vaccine against CanineCV.

CHAPTER 1
INTRODUCTION

Circoviruses are spherical, unenveloped viruses with a diameter of approximately 20 nm and a circular, single-stranded DNA genome of around 2 kb that are members of the Circoviridae family **(Todd et al., 2001)**. *Canine circovirus* (CanineCV) **(Anderson et al., 2017)** or *Dog Circovirus* (Dog CV) belongs to the genus *Circovirus*. They are the smallest known autonomously replicating, capsid-encoding animal pathogens. In the dog circovirus genome, two major (and opposing) transcriptional units contain 2 genes encoding (Exons), a replicase protein (Rep), as well as the capsid (Cap) proteins **(Kotsias et al., 2019)**. During the S phase of cell division, this virus replicates its genomes via a circular, ds replicative form DNA intermediate created by host cell DNA polymerases. The RF acts as a template for the production of viral ssDNA, which is most likely accomplished by the rolling circle replication process. Their genome consists of two coding and two non-coding parts. For unique viral replicase and capsid proteins, there are only 2 exons. Virion replicase protein has 303 amino acids while the capsid protein has 270 amino acids. A series of 30 Arginine amino acids from the amino terminus is used in the gene coding for the capsid. For DNA binding, this particular stretch is hypothesized to be important. **(Cruz et al., 2020, Sun et al., 2020)**. In the United States and Italy, Dogs with angiitis as well as gastroenteritis have been discovered to have canine circovirus **(Decaro et al., 2014, Li et al., 2013)**. Irrespective of the frequency of circovirus transmission, the clinical importance of canine circovirus in polymicrobial illnesses, and the distribution of this virion in those other wild animals, remains unclear **(Zaccaria et al., 2016)**. In the United Kingdom, a nearly similar circovirus was found in coyotes having meningoencephalitis. Currently, Canine circovirus disease pathogenesis remains unknown **(Plog et al., 2017)**. CanineCV has only been detected in the US **(Li et al., 2013, Kapoor et al., 2012)**, Italy **(Decaro et al., 2014)**, Germany (NCBI Acc No. KF887949), China (NCBI Acc No.KT946839) **(Hsu et al., 2016)**. Song *et al*. isolated Porcine Circovirus Type 2 in Racoon Dogs (Mangut) in China which caused failure in the reproduction of raccoon dogs and concluded the route of transmission of PCV2 from pigs to raccoon dogs **(Song et al., 2019)**. Canine circovirus (CanineCV) and Canine Parvovirus were discovered in samples from an epidemic of gastroenteritis in dogs in Argentina (South America) **(Kotsias et al. 2019)**. After a phylogenetic study, it was shown that the UBA-Baires virus is strongly related to virions from European strains **(Kotsias et al. 2019)**. As *Canine circovirus* can be fatal to the dog population, it is imperative to look for novel ways to combat

their infections. One such strategy can be developing vaccine candidates that can prevent dogs from *Canine circovirus* infections by generating a robust immune response. As a result, a multi-epitope vaccination against Canine circovirus was developed in this work using an *in-silico* strategy that targeted the virus's replicase and capsid proteins. As, these proteins are highly conserved, involved in viral pathogenesis, and can help to generate an immune response, they can be a good target for identifying epitopes that could be used to design a vaccine construct that can help protect the dog population from potentially lethal Canine circovirus infections. Such *in silico* studies have been previously performed to design vaccine constructs against Dengue**,** *Candida auris*, Human Cytomegalovirus, SARS-CoV-2, Lassa virus, Human Papillomavirus, Cervical cancer, and Hepatitis C Virus**.** In our study, various immunoinformatics tools were used to design a potential vaccine construct against *Canine circovirus*. First antigenic epitopes were selected and their interaction with dog leukocyte antigen (DLA) molecules was analyzed by docking and MD simulation studies. Then antigenic epitopes were linked with RS09 and flagellin adjuvants along with PADRE sequence by GGS linkers to construct a vaccine candidate. The physiochemical properties, antigenicity, allergenic potential, 2^{o} and 3^{o} structure of the designed vaccine construct were also predicted using different web servers. Finally, the ability of the vaccine construct to be cloned and expressed was also analyzed by the *in-silico* cloning method. This predicted vaccine has the potential to protect dogs from Canine circovirus infections. *Canine circovirus* (CanineCV) is a non-enveloped, covalently closed, circular single-stranded DNA virus belonging to the family Circoviridae. The genome of CanineCV consists of 2kb nucleotides containing two major open reading frames as follows: a. ORF1 (rep gene) encoding replicase for viral replication; b. ORF2(cap gene) encodes the capsid that participates in the host immune response. CanineCV is genetically close to porcine circovirus and is considered to be the first non-porcine mammalian circovirus first identified from the serum of dogs. Since the first evidence of CanineCV was reported, the virus has been identified in the serum, tissues, and fecal samples of dogs diagnosed with gastroenteritis and respiratory diseases in many countries **(Giraldo-Ramirez et al., 2020)**. The virus is also identified in the tissues of dogs with cases of angiitis, hemorrhage, and granulomatous lymphadenitis. Intriguingly, Canine CV is reported in healthy dogs from rural as well as urban areas. Further, there is evidence that suggests that CanineCV-diagnosed dogs having gastrointestinal problems are found

co-infected with other viral strains *viz.*, parvovirus, vesivirus, and astrovirus. Interestingly, CanineCV is found to act synergistically with other enteric viruses, for instance, parvovirus in developing gastrointestinal disorders in the dog. A recent report asserted that the CanineCV-infected dogs harboring Canine parvovirus-2 (CPV-2) have a high mortality rate compared to the dog infected with CPV-2 only. Immunosuppression and lymphoid depletion may occur in CanineCV-infected dogs as seen with other circoviral infections; however, its pathogenesis is poorly understood.

Apart from dogs, CanineCV is also found in the tissue samples collected from other canines including red fox, arctic fox, and wolves in Norway and Italy. The existence of CanineCV in wild animals implies an impending threat to the conservation of vulnerable wild species which may also act as a potential source of infection to dogs. Thus, it is imperative to search for novel avenues to control the spread of CanineCV infection in canines, especially dogs. With the recent progress in the field of bioinformatics, different strategies have been employed to design knowledge-based vaccines using the immunoinformatics approach. Multi-epitope vaccines (MEV) have been designed for Epstein-Barr and Chlamydia trachomatis viruses using immunoinformatics which is found to elicit effective humoral and cellular responses both *in vitro* and *in vivo*.

CHAPTER 2
REVIEW OF LITERATURE

Introduction

Canine circovirus (CanineCV) **(Anderson et al., 2017)** or Dog Circovirus (Dog CV) belongs to the genus Circovirus within the family Circoviridae. Canine CV has been isolated in 2012 as a part of genetic screening of canine samples for the new virus **(Kapoor et al., 2012)**. Since then various strains of CanineCV have been isolated in different regions of the globe including the USA, Italy, the UK, Croatia, China, Thailand, and Argentina.

Structure and genome organization

Canine circovirus is an icosahedral virus that is non-enveloped and has a circular, monomeric, ambisence, single-stranded DNA (ssDNA) genomic stretch of roughly 2000bp. They are the tiniest self-replicating, capsid-encoding animal pathogens ever discovered. In the dog circovirus genome, two major (and opposing) transcriptional units contain 2 genes encoding (Exons), a replicase protein (Rep), as well as the capsid (Cap) proteins **(Kotsias et al., 2019)**. Canine CV uses a circular, ds replicative form DNA intermediate created by host cell DNA polymerases during the S phase of cell division to replicate their genomes. The Replicative form acts as a template for viral ssDNA production, which is most likely accomplished by the rolling circle replication process.

Proteins

Canine circoviruses are a paragon of efficiency, replicating using the machinery of the host cell **(Todd., 2000)**. Due to the small number of proteins encoded in viral genomes, circoviruses and circovirus-like viruses have just two main proteins: (Rep) and (Cap). The most conserved circovirus protein is Rep, which possesses sequence patterns similar to proteins involved in rolling circle replication **(Ilyina and Koonin., 1992)**. Rep is involved in viral replication. Because it employs repetitive subunits to create the complete capsid structure of the virus, the capsid-associated protein (Cap) is highly diverse and is characterized by an N-terminal region rich in basic amino acids that may give DNA binding activity **(Niagro Niagro et al., 1998)**. The cap gene sequence has been used to divide Canine CV into two genotypes (Canine CV-1andCanine CV-2) **(Sun et al.,2019)**.

Epidemiology

In the United States and Italy, Dogs with angiitis as well as gastroenteritis have been discovered to have canine circovirus **(Decaro et al., 2014, Li et al., 2013)**. Irrespective of the frequency of circovirus transmission, the clinical importance of canine circovirus in polymicrobial illnesses, and the distribution of this virion in those other wild animals, remains unclear **(Zaccaria et al., 2016)**. In the United Kingdom, a nearly similar circovirus was found in coyotes having meningoencephalitis. Currently, Canine circovirus disease pathogenesis remains unknown (**Plog et al., 2017)**. CanineCV has only been detected in the US **(Li et al., 2013, Kapoor et al., 2012)**, Italy **(Decaro et al., 2014)**, Germany (NCBI Acc No.KF887949), China(NCBI Acc No.KT946839)**(Hsu et al., 2016)**. Song et al. discovered PorcineCircovirusType2 in Chinese Mangut Racoon Dogs (Mangut), which caused reproductive dysfunction in farmed raccoon dogs, and indicated that PCV2 may be transmitted from pigs to raccoon dogs **(Song et al., 2019)**. After a phylogenetic study, it was shown that the UBA-Baires virus is strongly related to virions from European strains **Kotsias et al., 2019. Zaccaria et al., 2016** used molecular technologies to test 389 internal organ samples from 277 domestic dogs and wild animals such as wolves, foxes, and badgers for the presence of dog circovirus (DogCV). They discovered that all DogCV-positive dogs had at least one other disease, such as canine distemper virus (CDV) and canine parvovirus type 2. They also sequenced the whole genome of DogCV and performed phylogenetic analysis. Canine circovirus strain NY214 was identified for the first time in dogs by **(Kapoor et al.,2012)**, who discovered that the strain is closely linked to swine circovirus. Canine circovirus genotype 1 (CaCV-1) was discovered in serum samples from several dogs (6 out of 205 animals examined), making it the first nonporcine circovirus to be verified to infect vertebrates. They looked at feces samples from diarrheal and healthy dogs and discovered that DogCV was extremely prevalent in diarrheal dogs in Taiwan. **Piewbang et al., 2018** used next-generation sequencing to identify CanineCV in samples from Thai dogs with respiratory symptoms. Natural recombination was found across distinct CanineCV lineages, according to genetic characterization and phylogenetic analysis of virtually full CanineCV genomes. **Li et al., 2013**, published the entire genome of a novel dog circovirus isolated from the liver of a dog with severe hemorrhagic gastroenteritis, angiitis, and granulomatous lymphadenitis. DogCV was also found in the feces of 19

out of 168 (i.e.11.3%) diarrhoea dogs and 14 out of 204 (i.e.6.9%) healthy dogs, as well as in the blood of 19 out of 409 (i.e.3.3%) dogs with thrombocytopenia and neutropenia, fever of uncertain etiology, or a recent tick bite. DogCV was found in the liver and spleen of a dog who died of histiocytic sarcoma, according to Plog *et al*., but it was not linked to the virus infection. They discovered that the strain (NCBI Acc. No. KT283604) is closely linked to UCD-1, UCD-2andBari, and hence belongs to DogCV's 'cluster 1'.

Symptoms and lesions

CanineCV has been related to a number of disorders, including vasculitis, haemorrhages, thrombocytopenia, neutropenia, and diarrhoea, although it may also be seen in healthy dogs. CanineCV-infected dogs are frequently co-infected with other enteric or respiratory diseases. As a result, the pathogenic function of CanineCV remains unknown. Canine circovirus has been linked to gastrointestinal problems in dogs. In the United States, Dogs with angiitis as well as gastroenteritis have been discovered to have canine circovirus **(Li et al., 2016)**. A probable link between CaniveCV and canine enteritis has been identified in recent years **(Li et al., 2016, Thiawong et al., 2016)**. CaniveCV has also been linked to necrotizing lymphadenitis and vasculitis in pigs, both of which are caused by porcine circovirus type2 infections **(Segalés, 2012)**. CaniveCV infection has been linked to hemorrhagic enteritis in dogs in previous studies, but there is no evidence on a direct link between the severity of diarrhoea and CaniveCV infections.

Prevention and Control

Supportive therapy, similar to that used to treat other intestinal illnesses, is recommended. Because circovirus is linked to vasculitis, dogs should be closely examined for signs of disseminated intravascular coagulation and cardiovascular impairment.

Review of Literature & research gap identification

Viruses' codon use patterns represent a sequence of evolutionary modifications that allow them to adjust their survival rates and fitness concerning the external environment and, most significantly, their hosts. Nucleotide triplets encode the amino acid sequence of proteins. The conventional genetic code, which consists of 61 sense codons that are translated into 20 amino acids, is used by the vast majority of species. As a result, most amino acids are encoded by several synonymous codons that are not

uniformly distributed. In various species, this codon usage bias might have different origins and implications. CUB is thought to be caused by two key factors translational (selection) and mutational pressure (which is a bit of a misnomer) **(Bulmer., 1987)**. Translational pressure refers to a preference for codons that are best suited for translation in a particular situation. Different probabilities of different substitution types cause mutational pressure. The GC content, or the sum of the relative abundance of G and C nucleotides, is thought to be a major determinant of codon use **(Chen et al., 2004, Sharp et al.,1993**). Deoxycytidine methylation and subsequent deamination in the CpG (C-phosphate-G) dinucleotide causes in C-T substitution are two more processes that may contribute to mutational pressure **(Lister et al., 2009)**.

Most amino acids can be translated by more than one codon, making the genetic code redundant. This redundancy is an important aspect in protein manufacturing efficiency and precision, as well as maintaining the protein's amino-acid sequence. Synonymous codons are alternative codons within the same cluster that code the same amino acid, even though their corresponding tRNAs may differ in relative abundance in cells and ribosome recognition speed. Within and across genomes, however, synonymous codons just aren't selected at random, a condition referred as "codon usage bias" **(Grantham et al., 1980, Martin et al., 1989)**. This phenomenon has been seen in a wide spectrum of taxa, from prokaryotes to eukaryotes and viruses. Several variables have been recognised as impact on codon usage patterns, including mutation pressure, translational selection, secondary protein structure, replication selective transcription, protein surface chemistry and solubility in water, and the physical factors **(Gu et al., 2004, Liu et al., 2011, Ma et al., 2013, Moratorio et al., 2013, Sharp et al., 1988, Tao et al., 2009)**. The results are startling when the size of the viral genome and other viral properties, such as its dependency on the host apparatus for crucial physiological functions, are compared to those of prokaryotic and eukaryotic genomes. The interaction between the virus's codon usage and that of its host is expected to have an impact on the virus's overall survival, fitness, immune system evasion, and evolution **(Shackelton et al., 2006)**. As a result, understanding how viruses use codons can help us learn more about their molecular evolution, and enhance vaccine design and knowledge of viral gene control, as effective production of viral proteins may be essential to produce immunogenicity.

Jenkins and Holmes., 2003 studied codon usage bias in a wide range of genetically and ecologically diverse human RNA viruses. Their analysis showed that the overall extent of codon usage bias in RNA viruses is low and that there is little variation in bias between genes. Furthermore, the strong correlation between base and dinucleotide composition and codon usage bias suggested that mutation pressure rather than natural (translational) selection is the most important determinant of the codon bias observed. They also detected correlations between codon usage bias and some characteristics of viral genome structure and ecology, with increased bias in segmented and aerosol-transmitted viruses and decreased bias in vector-borne viruses. This suggests that translational selection may also have some influence in shaping codon usage bias.

Coleman et.al., 2008 synthesized de novo large DNA molecules using hundreds of over-or underrepresented synonymous codon pairs to encode the poliovirus capsid protein. Underrepresented codon pairs caused decreased rates of protein translation, and polioviruses containing such amino acid-independent changes were attenuated in mice. Polioviruses thus customized were used to immunize mice and provided protective immunity after the challenge. This "death by a thousand cuts" strategy could be generally applicable to attenuating many kinds of viruses.

Xu et al., 2008 performed an analysis of synonymous codon usage and evolution of begomoviruses and found that synonymous codon usage variations in the protein-coding genes of begomoviruses are mainly influenced by mutation bias. Base composition analysis suggested that the codon usage bias of *AV1* and *BV1* genes is significant and their expressions are high. Fourteen codons were determined as translational optimal ones according to the comparison of codon usage patterns between highly and lowly expressed genes. Interestingly the codon usages between begomoviruses from the Old and the New Worlds are apparently different, which supports the idea that the bipartite begomoviruses of the New World might originate from bipartite ones of the Old World, whereas the latter evolved from the Old World monopartite begomoviruses.

Tao et al., 2009 performed an Analysis of synonymous codon usage in classical swine fever virus (CSFV). The general correlation between base composition and codon usage bias suggests that mutational pressure rather than natural selection is the

main factor that determines the codon usage bias in CSFV. They also observed that the relative abundance of dinucleotides in CSFV is independent of the overall base composition but is still the result of differential mutational pressure, which also shapes codon usage. In addition, other factors, such as the subgenotypes and aromaticity, also influence the codon usage variation among the genomes of CSFV.

Fu., 2010 performed Codon usage bias in herpesvirus. The values of the effective number of codons revealed that the majority of the herpesviruses did not have high codon bias, with the exceptions of only simplex viruses and some varicella viruses. The results of correspondence analysis, a plot of ENC, and GC content at the third position of the codon analysis of preferred codons all revealed that the codon bias was mainly influenced by nucleotide composition. This suggests that mutation pressure is the main factor influencing codon usage bias in herpesviruses.

Zhang et al., 2011 performed an Analysis of codon usage and nucleotide composition bias in polioviruses. Their result shows that the overall extent of codon usage bias in poliovirus samples is low (mean ENC = 53.754 > 40). The general correlation between base composition and codon usage bias suggests that mutational pressure rather than natural selection is the main factor that determines the codon usage bias in those polioviruses. Depending on the RSCU data, it was found that there was a significant variation in bias of codon usage among the three genotypes. Geographic factor also has some effect on the codon usage pattern (exists in the genotype-1 of polioviruses). No significant effect in gene length or vaccine-derived polioviruses (DVPVs), wild viruses, and live attenuated viruses was observed on the variations of synonymous codon usage in the virus genes.

Liu et al., 2012 studied patterns and influencing factors of synonymous codon usage in Porcine Circovirus. They carried out a comprehensive analysis of codon usage patterns in the PCV genome, by calculating relative synonymous codon usage (RSCU), effective number of codons (ENC), dinucleotides, and nucleic acid content of the PCV genome. They found that PCV genomes have a relatively much lower content of GC and codon preference, this result shows that nucleotide constraints have a major impact on its synonymous codon usage. The results of the correspondence analysis indicate codon usage patterns of PCV of various genotypes,

various sub-genotypes changed greatly, and significant differences in codon usage patterns of each virus of Circoviridae. There is much comparability between PCV and its host in their synonymous codon usage, suggesting that the natural selection pressure from the host factor also affects the codon usage patterns of PCV. In particular, PCV genotype II is in synonymous codon usage more similar to the pig than to PCV genotype I, which may be one of the most important molecular mechanisms of PCV genotype II to cause disease. The calculation results of the relative abundance of dinucleotides indicate that the composition of dinucleotides also plays a key role in the variation found in synonymous codon usage in PCV. The results of their studies suggest that synonymous codon usage patterns of PCV genome are the result of the interaction between mutation pressure and natural selection from its host. The information from this study may not only have theoretical value in understanding the characteristics of synonymous codon usage in PCV genomes but also have significant value for the molecular evolution of PCV.

Belalov and Lukashev.,2013 analyze the relative impact of mutational pressure components on codon usage bias in RNA viruses. They interpreted that total GC content was a poor descriptor of viral genome composition and causes of codon usage bias and genomic nucleotide content was the single most important factor of synonymous codon usage. Moreover, the choice between compatible amino acids (e.g., leucine and isoleucine) was strongly affected by genomic nucleotide composition. Dinucleotide composition at codon positions 2-3 had an additional effect on codon usage. Together with mononucleotide composition bias, it could explain almost the entire codon usage bias in RNA viruses. On the other hand, strong dinucleotide content bias at codon position 3-1 found in some viruses had very little effect on codon usage.

Du et al., 2014 studied the selection on synonymous codons in mammalian rhodopsins andits possible role in optimizing translational processes. They performed phylogenetic codon-based likelihood models to explore patterns of codon usage bias in a dataset of 18 mammalian rhodopsin sequences, the protein mediating the first step in vision in the eye, and one of the most highly expressed genes in vertebrates. Their results show patterns ofselection in mammalian rhodopsin appear to be correlated with post-transcriptional and translational processes. They found significant evidence for selection at synonymous sites using phylogenetic mutation-

selection likelihood models, with C-ending codons found to have the highest relative fitness and to be significantly more abundant at conserved sites. They also found significant differences in codon usage bias between rhodopsin loops versus helices, though there was no significant difference in the mean synonymous substitution rate between these motifs.

Chen et.al., 2014 performed the analysis of synonymous codon usage in Porcine Circovirusto investigate the codon usage bias of PCV, and the genomic sequences of PCV1 and PCV2. Their results showed that the codon usage bias of PCV was very low. An effective number of codons (ENC) plot analysis indicated that mutational pressure influences the codon usage bias of PCV. Neutrality plot analysis showed that mutation bias dominated natural selection in shaping the codon usage bias of PCV1, but mutation bias and natural selection contributed equally to the codon usage bias of PCV2. Principal component analysis showed that different ORFs and dinucleotide patterns were also factors influencing the codon usage bias of PCV.

Aravind et al., 2014 performed a Bioinformatics Study involving the Characterization of Synonymous Codon Usage Bias in the Duck Enteritis Virus Glycoprotein D (gD) Gene. Their study aimed to identify the codon usage bias between the newly identified Duck Enteritis Virus (DEV) gD gene (GenBank accession No. KC915041) and the gD-like gene of 23 other reference herpesviruses. Their results showed that the codon of gD gene of DEV was having a strong bias towards the synonymous codons with A and T at the third codon position. A high level of diversity in codon usage bias existed, and the effective number of codons used in a gene plot revealed that the genetic heterogeneity in gD gene of herpesviruses was constrained by the G+C content. The phylogenetic analysis suggested that DEV was evolutionarily closer to Alpha herpesvirinae, there was no significant deviation in codon usage in different virus strains. 17 codons were showing distinct usage differences between DEV and Escherichia coli, 22 between DEV and Homo sapiens but only 15 codons between DEV and yeast. Therefore, the yeast expression system may be more suitable for the expression of DEV genes. The results were encouraging regarding bioinformatics data of an economically important poultry pathogen and could be highly useful for the development of new-generation vaccines, diagnostics, and studying the evolution of the duck enteritis virus.

Roy et al., 2015 predicted codon and amino acid usage patterns in *Bifidobacterium* genus, one of which is associated with probiotic activities and is present in gut of human,animals, and insects. In their work, they anticipated that GC content is high in the genome of *Bifidobacterium* which is around 60.48%, so GC-rich codon usage was high in the genomic content. This also makes the prediction that codons of GC richness are preferred over AT richness. Also, Cytosine has more preference at the third wobble position. The results have also been shown graphically by a heat map and rose plots. Results of the heat map also predicted the presence of amino acids such as Alanine, Glycine, Leucine, Valine, Aspartic acid, and Arginine in high frequency in the genome showing high usage of these amino acids. The Neutrality Plot constructed also predicted that ribosomal proteins have a high level of expression during cell division and growth. Finally, they described from their study that evolutionary selection is one of the important aspects governing the codon usage bias in *Bifidobacterium* genus.

Soria-Guerra et al., 2015 presented a comprehensive compilation of the most advantageous online immunological software and searchable, to facilitate the design and development of vaccines. An outlook on how these tools are supporting vaccine development has been presented. HIV and influenza have been taken as examples of promising developments in vaccination against hypervariable viruses.

Fan et al., 2015 introduced over 300 silent mutations into the genome of a seasonal H1N1 influenza virus. The resultant mutant was significantly attenuated in mammalian cells and mice, yet it grew well in embryonated eggs. A single dose of intranasal vaccination-induced potent innate, humoral, and cellular immune responses, and the mutant could protect mice against homologous and heterologous viral challenges. The attenuated mutant could also be used as a vaccine master donor strain by introducing hemagglutinin and neuraminidase genes derived from other strains. Thus, their approach is a successful strategy to generate attenuated viruses for future applications as vaccines.

Baker et al., 2015 reviewed the utility of codon deoptimization to alter the replicative fitness of Influenza A viruses. Vaccines based on changes in codons could principally

be generated rapidly in response to any emerging, potentially pandemic, influenza viruses. To date, all studies of deliberate virus gene recoding have employed two specific approaches, codon usage bias or codon-pair bias, which have resulted in cell type-specific virus fitness loss *in vitro*, and attenuation *in vivo*. Importantly, codon deoptimized viruses will express an intact repertoire of antigenic sites because the amino acid sequence is not altered, thus conserving maximum immunogenicity for cellular and humoral immunity. Using new technologies, viral segments can be produced by *de novo* synthesis in a few weeks, hastening the time from surveillance of seasonal or pandemic strains to vaccine development.

Liao et al., 2015 studied recurrent positive selection and heterogeneous codon usage bias events leading to the coexistence of divergent pigeon circoviruses. The capsid genes from 14 pigeon circovirus (PiCV) sequences, collected from Taiwan between 2009 and 2010, were sequenced and compared with 14 PiCV capsid gene sequences from GenBank. Based on the pairwise comparison, PiCV strains from Taiwan shared 73.9-100 % nucleotide identity and 72-100 % amino acid identity with those of the 14 reported PiCV sequences. Phylogenetic analyses revealed that Taiwanese PiCV isolates can be grouped into two clades: clade 1 comprising isolates from Belgium, Australia, USA, Italy, and China, and clade 2 showing close relation to isolates from Germany and France. Recurrent positive selection was detected in clade 1 PiCV lineages, which may contribute to the diversification of predominant PiCV sequences in Taiwan. Further observations suggest that synonymous codon usage variations between PiCV clade 1 and clade 2 may reflect the adaptive divergence in the translation efficiency of capsid genes in infectious hosts. Variation in selective pressures acting on the evolutionary divergence and codon usage bias of both clades explains the regional coexistence of virus sequences congeners prevented from competitive exclusion within an island such as Taiwan. Genotyping results also provide insight into the aetiological agents of PiCV outbreak in Taiwan and we present a comparative analysis of the central coding region of PiCV genome. From the sequence comparison results of 28 PiCVs which differ in geographical origin and columbidae species, they identified conserved regions within the capsid gene that are likely to be suitable for primer selection and vaccine development.

Nasrullah et al., 2015 performed a comprehensive analysis of synonymous codon usage patterns in Marburg virus genomes. Multiple codon analysis approaches and statistical methods were performed to determine overall codon usage patterns, biases in codon usage, and the influence of various factors, including mutation pressure, natural selection, and its two hosts, Homo sapiens and Rousettus aegyptiacus. Nucleotide composition and relative synonymous codon usage (RSCU) analysis revealed that MARV shows mutation bias and prefers U- and A-ended codons to code amino acids. Effective number of codons analysis indicated that overall codon usage among MARV genomes is slightly biased. The Parity Rule2 plot analysis showed that GC and AU nucleotides were not used proportionally which accounts for the presence of natural selection. Codon usage patterns of MARV were also found to be influenced by its hosts. Their result indicates that MARV has evolved codon usage patterns that are specific to both of its hosts.

Xu et al., 2015 performed analysis of synonymous codon usage patterns in duck Circovirus. In their study, relative synonymous codon usage (RSCU) values, nucleotide contents, and an effective number of codon (ENC) values were calculated and compared among open reading frames (ORFs) of 53 DuCV genomes. Their results reveal that most of the codons ended with C and the overall bias is not remarkable in DuCV. A comparative analysis of codon contents and ENC values indicated that mutation pressure is the most significant factor responsible for the evolutional processes of codon usage bias in DuCV. However, other factors, such as composition constraints, translation selection, hydrophobicity, and aromaticity should not be ignored. Finally, principal component analysis (PCA) and hierarchical clustering method were performed based on RSCU. A significant difference of codon usage bias exists in DuCV-1 and DuCV-2 genotypes, but codon usage pattern of DuCV from the different epidemic areas or subtypes fails to influence the formation of codon usage bias. Analysis of the relationship of synonymous codon usage variation based on the two genotypes suggests that DuCV-2 is more conservative than DuCV-1, which may be because of recombination events. Moreover, there are distinct differences in the degree of codon usage pattern evolution in different function genes, rep and cap. Therefore, the genotypes, subtypes, and different functional genes also relate to the pattern of synonymous codon usage.

Xiang et.al., 2015 performed a comparative analysis of Codon Usage Bias Patterns in Microsporidian Genomes between eight taxonomically distinct microsporidian genomes: Encephalitozoon intestinalis, Encephalitozoon cuniculi, Spraguea lophii, Trachipleistophora hominis, Enterocytozoon bieneusi, Nematocida parisii, Nosema bombycis, and Nosema ceranae. While the CUB was found to be weak in all eight Microsporidia, nearly all (98%) of the optimal codons in S. lophii, T. hominis, E. bieneusi, N. parisii, N. bombycis and N. ceranae are fond of A/U in the third position whereas most (64.6%) optimal codons in the Encephalitozoon species E. intestinalis and E. cuniculi are biased towards G/C. Although nucleotide composition biases are likely the main factor driving the CUB in Microsporidia according to correlation analyses, directed mutational pressure also likely affects the CUB as suggested by ENc-plots, correspondence, and neutrality analyses. Overall, the Encephalitozoon genomes were found to be markedly different from the other microsporidians and, despite being the first sequenced representatives of this lineage, are uncharacteristic of the group as a whole.

Kumar et al., 2016 examined the codon usage bias of Equine influenza viruses (EIVs) of H3N8 subtype, causing severe respiratory infections in horses, and are responsible for significant outbreaks worldwide. In this study, they have evaluated codon patterns of genome of Equine influenza virus in 92 strains by computing several codon usage indices such as Relative synonymous codon usage and results have shown a preference for codons ending with A/U. Many statistical analyses were also performed showing mutational and natural selection both factors affected host-virus interaction, it also predicted that these viruses are less adapted to host horse. This study may assist the researchers in increasing their understanding of factors involved in Equine influenza viral adaptation, evolution, and fitness towards their hosts.

Butt et al., 2016 examined codon usage bias in Zika virus by using various computational tools and software including CodonW, SPSS23, CAIcal server, eRCDI server, Recombination Detection Program server. They have also interpreted the evolution of codon usage patterns of viral proteins by using MEGA7 software as variation in codon preference to assist the virus to adapt itself to the host environment. In their report, they predicted that natural selection is favoured over

mutational for codon bias in Zika virus. Prediction of host adaptation of Zika virus to optimize themselves to changing external environment allowing the virus to survive inside host was done for the first time providing insight of survival of Zika virus under multiple hosts. These investigations are concerned to be of great advantage while evaluating preventive strategies for Zika virus.

Wang et al., 2016 performed a comprehensive analysis on the codon usage pattern in46 ZIKV strains by calculating the effective number of codons (ENc), codon adaptation index (CAI), relative synonymous codon usage (RSCU), and other indicators. The results indicate that the codon usage bias of ZIKV is relatively low. Several lines of evidence support the hypothesis that translational selection plays a role in shaping the codon usage pattern of ZIKV. The results from a correspondence analysis (CA) indicate that other factors, such as base composition, aromaticity, and hydrophobicity may also be involved in shaping the codon usage pattern of ZIKV. Additionally, the results from a comparative analysis of RSCU between ZIKV and its hosts suggest that ZIKV tends to evolve codon usage patterns that are comparable to those of its hosts. Moreover, selection pressure from *Homo sapiens* on the ZIKV RSCU patterns was found to be dominant compared with that from *Aedes aegypti* and *Aedes albopictus*. Their findings contribute to understanding the evolution of ZIKV and its adaption to its hosts.

Liu et al., 2016 performed a Genome-Wide Analysis of the Synonymous Codon Usage Patterns in Riemerella anatipestifer isolates from the available 12 sequenced genomes by multiple codon and statistical analysis. Nucleotide compositions and relative synonymous codon usage (RSCU) analysis revealed that A or U ending codons are predominant in RA. Neutrality analysis found no significant correlation between GC12 and GC3 ($p > 0.05$). Correspondence analysis and ENc-plot results showed that natural selection dominated over mutation in the codon usage bias. The tree of cluster analysis based on RSCU was concordant with the dendrogram based on genomic BLAST by the neighbor-joining method. By comparative analysis, about 50 highly expressed genes that were orthologs across all 12 strains were found in the top 5% of high CAI values. Based on these CAI values, they concluded that RA contains several predicted highly expressed coding sequences, involved in transcriptional regulation and metabolism, reflecting their requirement for dealing

with diverse environmental conditions. These results provide some useful information on the mechanisms that contribute to codon usage bias and evolution of RA.

Dash et al., 2017 designed an epitope-based peptide vaccine against EBOV, using a combination of B-cell and T-cell epitope predictions, followed by molecular docking and molecular dynamics simulation approach. In their study, protein sequences of all glycoproteins of EBOV were collected and examined via in silico methods to determine the most immunogenic protein. From the identified antigenic protein, the peptide region ranging from 186 to 220 and the sequence HKEGAFFLY from the positions of 154–162 were considered the most potential B-cell and T-cell epitopes, correspondingly. The peptide (HKEGAFFLY) interacted with HLA-A*32:15 with the highest binding energy and stability, and also a good conservancy of 83.85% with maximum population coverage. Their study has shown that the designed epitopes could manifest vigorous enduring defensive immunityagainst EBOV**.**

Roy et al., 2017 examined the effect of codon usage and amino acid usage in HIV (human immunodeficiency virus) and also discussed the influence of human host machinery on the progression of disease caused by viruses, as they have discussed the adaptation of virus to host by comparative analysis of codon usage in both host and virus. Sequences of envelope gene which evolve fastest in HIV-1 virus were retrieved from HIV database and sequences were aligned by using clustal omega and further MEGA software was used for phylogenetic analysis. Various statistical parameters such as Relative Synonymous Codon Usage, Similarity Index, Correspondence analysis, and evolutionary rate were examined by using the software CodonW and SWAKK server to predict the occurrence of variable patterns among codon and amino acid. These parameters also conclude that natural selection as a main cause of variation among patterns. The study concluded that LTNP (Long term non- progressor) gene sequences of HIV-1 have a stronger pure selection as compared to RP (rapid progressor) genes and have stronger functional constraints. Therefore, viral strains from LNTP are less evolved and are less capable of escaping host immunological responses as compared to RP genes.

Kumar and Kumar., 2017 inspected the codon usage bias of the genome of Newcastle virus, pathogenic to both avian and non-avian species. Total of 186 polymerase genes comprising nucleoprotein (N), phosphoprotein (P), and large polymerase (L) proteins of Newcastle Virus retrieved from GenBank were examined for variation in synonymous codon patterns. Relative synonymous codon usage and effective number of codons, correspondence analysis were some of the statistical parameters evaluated by using Codon W 1.4.4 to explore the codon usage bias and concluded from the results that mutational selection is the major factor causing codon bias in Newcastle virus genome and also aromaticity is one of the other factor affecting the codon bias.

Lee et al., 2017 studied the *Lactobacillus salivarius*, bacterium species present in animal gut and is one of the promising gut probiotic bacterium. In this study firstly isolation and then sequencing of draft sequences was performed by using illumina sequencing method of 21 genomes of *L.salivarius*, which include the genomes from human, pigs, and chickens. Then, sequences were used further to predict the adaptation of the genome to their host and phylogenetic analysis of genomes was also completed by using MEGA7 software. This analysis explained that 56 protein coding genes for extracellular proteins and 124 orthologs (generated by using GASSST 1.28) that are related to the production of exopolysaccharides are mutable and have revealed that two factors are responsible for the host adaptation of bacteria one is by the ability of bacteria to gain niche adhesion and by utilization of efficient nutrients. This study provides insight of genome of *L.salivarius* and other mutualistic bacteria.

Castells et al., 2017 performed a Genome-wide analysis of codon usage bias in Bovine Coronavirus and concluded that the global codon usage among BCoV strains is similar. Significant differences of codon preferences in BCoV genes with codon usage of Bos taurus host genes were found. Most of the highly frequent codons are U-ending. G + C compositional constraint and dinucleotide composition also play a role in the overall pattern of BCoV codon usage. The results of these studies revealed that mutational bias is a leading force shaping codon usage in this virus. Additionally, relative dinucleotide frequencies, geographical distribution, and evolutionary processes also influenced the codon usage pattern.

Athey et al., 2017 performed a new and updated resource for codon usage tables quantifying codon usage bias for different organisms and found that is useful not only for codon optimization, but also for evolutionary and translation studies: phylogenetic relations of organisms, and host-pathogen co-evolution relationships, may be explored through their codon usage similarities. Furthermore, codon usage has been shown to affect protein structure and function by interfering with translation kinetics, and cotranslational protein folding.

Gun et al., 2018 performed a Comprehensive Analysis and Comparison on the Codon Usage Pattern of the Whole *Mycobacterium tuberculosis* Coding Genome. 12 *Mycobacterium tuberculosis* genomes from the different *areas* were downloaded from the GeneBank. The correlations between G3, GC12, whole GC content, codon adaptation index, codon bias index, and so on of *Mycobacterium tuberculosis* genomes were calculated. The ENC-plot, relationship between A3/(A3 +T3) and G3/(G3 +C3), GC12 versus GC3 plot, and the RSCU of overall/separated genomes all showed that the codon usage bias exists in all 12 *Mycobacterium tuberculosis* genomes. The relationship between CBI and the equalization of ENC shows a strong negative correlation between them. The relationship between protein length and GC content (GC3 and GC12) shows that more obvious differences in the GC content may be in shorter protein. Their results show that codon usage bias existing in the *Mycobacterium tuberculosis* genomes could be used for further study on their evolutionary phenomenon.

Galtier et al., 2018 performed Codon Usage Bias in 30 non-model species of animals, each from a different family, covering a wide range of effective population sizes. They disentangled the effects of translational selection and GC-biased gene conversion on codon usage by separately analyzing GC-conservative and GC-changing mutations. They report evidence for effective translational selection on codon usage in large-Ne species of animals, but not in small-Ne ones, in agreement with the nearly neutral theory of molecular evolution. C- and T-ending codons tend to be preferred over synonymous G- and A-ending ones, for reasons that remain to be determined. They uncovered a conspicuous effect of GC-biased gene conversion, which is widespread in animals and the main force determining the fate of AT & GC mutations.

Franzo et al., 2018 performed the analysis of genome composition and codon bias and found distinctive patterns between avian and mammalian circoviruses which suggest a potential recombinant origin for *Porcine circovirus 3.* Analysis of several genome compositions and codon bias parameters of circoviruses infecting avian and mammalian species demonstrated the presence of quite distinctive patterns between the two groups. A higher deviation from the expected values based only on mutational patterns was observed for mammalian circoviruses both at dinucleotide and codon levels. A stronger selective pressure was estimated to shape the genome of mammalian circoviruses, particularly in the *Cap* encoding gene, compared to avian circoviruses. These differences could be attributed to different physiological and immunological features of the two host classes and suggest a trade-off between a tendency to optimize the capsid protein translation while minimizing the recognition of the genome and the transcript molecules. The recently identified *Porcine circovirus 3* (PCV-3) had an intermediate pattern in terms of genome composition and codon bias particularly, its *Rep* gene appeared closely related to other mammalian circoviruses (especially bat circoviruses) while the *Cap* gene more closely resembled avian circoviruses. These evidences, coupled with the highly selective forces modelling the PCV-3 *Cap*gene composition, suggest the potential recombinant origin, followed or preceded by a host jump, of this virus.

Karumathil et al., 2018 performed Synonymous Codon Usage Bias in West African and Central African Strains of Monkeypox Virus. Various trends associated with synonymous codon usage in chosen monkeypox viral genomes were investigated, and the results were reported. Identification of factors that influence codon usage in chosen monkeypox viral genomes was done using various codon usage indices, such as the relative synonymous codon usage, the effective number of codons, and the codon adaptation index. The Spearman rank correlation analysis and a correspondence analysis were used for correlating various factors with codon usage. Their results revealed that mutational pressure due to compositional constraints, gene expression level, and selection at the codon level for utilization of putative optimal codons are major factors influencing synonymous codon usage bias in monkeypox viral genomes. A cluster analysis of relative synonymous codon usage values revealed a grouping of more virulent strains as one major cluster (Central African strains) and a grouping of less virulent strains (West African strains) as another major

cluster,indicating a relationship between virulence and synonymous codon usage bias. Their study concluded that a balance between the mutational pressure acting at the base composition level and the selection pressure acting at the amino acid level frames synonymous codon usage bias in the chosen monkeypox viruses. The natural selection from the host does not seem to have influenced the synonymous codon usage bias in the analyzed monkeypox viral genomes.

Li et.al. 2018 performed a deep codon usage analysis of porcine circovirus 3 (PCV3). PCV3 sequences were classified into two clades: PCV3a and PCV3b, confirmed by principal component analysis. Additionally, the degree of codon usage bias of PCV3 was slightly low as inferred from the analysis of the effective number of codons. The codon usage pattern was mainly affected by natural selection, but there was a coeffect of mutation pressure and dinucleotide frequency. Moreover, based on similarity index analysis, codon adaptation index analysis, and relative codon deoptimization index analysis, they found that PCV3 might pose a potential risk to public health though with unknown pathogenicity. Their work reinforces the systematic understanding of the evolution of PCV3, which was reflected by thecodon usage patterns and fitness of this novel emergent virus.

Yao et al., 2019 studied the codon usage pattern of *Flaviridae* viruses, causing infection in humans and transmitted from mosquitos, and sandflies. In their work, Yao andcolleagues examined the evolutionary pattern of the virus and provide insight into the prediction of medication against the sites which are involved in causing infection and also involved in host-parasite interactions. Overall results interpreted a low level of codon bias and a high level of GC content in the genome of 65 virus strains under study. Results have also shown favourers of A or U at the third site of codon and all the statistical analyses such as neutrality plot, correlation analysis and ENC values (effective number of codons) performed computationally have shown a preference for natural selection over mutational for codon usage bias. Phylogenetic analysis finally concluded that codon usage bias as one of the major evolutionary aspects to study the evolution of *Flaviridae* viruses.

Khandia et al., 2019 analysed the codon usage and adaptation of Nipah virus and also the adaptation of viruses to their host. As the Nipah virus is becoming a huge

threat among people due to its high rate of transmission from human to human brings the need to understand the insights of Nipah virus by examining the codon usage, and amino acid usage bias of genome by using CodonW software. Further, factors influencing the adaptation of virus to host machinery for protein biogenesis were also examined. Host adaptation of Nipah virus was also compared with the host adaptation of *Hepavirus* genus among 10 hosts. Overall very low amount of bias was observed in Nipah virus codon usage and also aromaticity, one of thefactors of codon bias has no impact on the genome of Nipah virus. Prediction of preferred codon patterns in Nipah virus genome was compared with tRNA pool of human host. Statistical parameters such as codons adaptation index, similarity index, tRNA adaptation index were examined to predict the host adaptation of Nipah virus inferring that natural selection is the major parameter for host adaptation as compare to mutational parameter. Also, African green monkeys have been predicted as suitable host for Nipah virus adaptation as compared to others.

Li et al., 2019 studied torque teno sus virus, the causative agent of porcine circovirus associated disease (PCVAD). Codon usage analysis of ORF1 (open reading frame) of virus encoding viral capsid protein was performed and further phylogenetic analysis was also examined to explore the evolutionary history of the proteins of the virus. Results examined the preference of ATs as compared to GCs in the genome and natural selection is the proposed cause of codon bias in the viral species. Codons ending with A- were favoured and these results were interpreted from computationally computed statistical parameters such as Codon Adaptive Index, Relative Synonymous Codon Usage, Relative Codon Adaptation Index, and similarity index. *Sus scrofa* was examined to be the preferred host over others. This information provides insight for genome of virus and its preventive measures.

Baha et al., 2019 studied a Comprehensive analysis of the genetic and evolutionary features of the hepatitis E virus. Sequences of HEV strains isolated between 1982 and 2017 were retrieved and multiple analyses were performed to determine overall codon usage patterns, effects of natural selection and/or mutation pressure, and host influence on the evolution of HEV ORFs. Analysis was performed to estimate the spatial-temporal evolution of HEV. The results indicated an A/C nucleotide bias and ORF-dependent codon usage bias affected mainly by natural selection.

Biswas et al., 2019 performed Codon Usage Bias Analysis of Citrus tristeza virus. The codon usage patterns of the coat protein (CP) genes of 122 CTV isolates originating from three economically important citrus hosts (55 isolate from Citrus sinensis, 38 from C. reticulata, and 29 from C. aurantifolia) were studied using several codon usage indices and multivariate statistical methods. The present study shows that CTV displays low codon usage bias (CUB) and higher genomic stability. Neutrality plot and relative synonymous codon usage analyses revealed that the overall influence of natural selection was more profound than that of mutation pressure in shaping the CUB of CTV. In one of the novel research, the group designed two artificial polyepitope T-cell immunogens, one of which (EV.CTL) includes cytotoxic and the other (EV.Th) – T helper epitopes of Ebola virus proteins using original TEpredict/PolyCTL Designer software. Synthesized genes were cloned in pcDNA3.1 plasmid vector. Target gene expression was estimated by the synthesis of specific mRNAs and proteins in cells transfected with recombinant plasmids. The immunogenicity of obtained DNA vaccine constructs was evaluated according to their capacity to induce a T-cell response in BALB/c mice using IFNγ ELISpot and ICS.

Tsai et al., 2019 performed Phylogeographic and genetic characterization of porcine circovirus type 2 in Taiwan from 2001-2017. They confirmed PCV2d-2 as the predominant PCV2 genotype in Taiwan. The introduction of PCV2 vaccination was indeed effective for PCV2 control in the field. However, the fast evolution rate of PCV2 may have led to the observed genotypic shift of this virus in Taiwan. Their study provides a foundation for researchers to consider new PCV2 vaccine design strategies and provide several directions for future study.

Krasovec and Filatop., 2019 studied Codon Usage Bias in Diatoms belonging to 35 genera from 4 classes. They report that most of the diatom species studied have surprisingly modest CUB (mean Effective Number of Codons, ENC = 56), with some exceptions showing stronger codon bias (ENC = 44). Modest codon bias in most studied diatom species may reflect the extreme disparity between astronomically large census and modest effective population size (Ne), with fluctuations in population size and linked selection limiting long-term Ne and rendering selection for optimal codons less efficient.

AUTHOR	TITLE	SOURCE	FINDINGS
Jenkins and Holmes., 2003	The extent of codon usage bias in human RNA viruses and its evolutionary origin	*Virus research*	translational selection has some influence in shaping codon usage bias
Coleman et.al., 2008	Virus attenuation by genome-scale changes in codon pair bias	*Science*	synthesized de novo large DNA molecules using hundreds of over- or underrepresented synonymous codon pairs to encode the poliovirus capsid protein
Xu et al., 2008	Analysis of synonymous codon usage and evolution of begomoviruses	*Journal of Zhejiang University Science B*	performed an analysis of synonymous codon usage and evolution of begomoviruses and found that synonymous codon usage variations in the protein-coding genes of begomoviruses are mainly influenced by mutation bias
Tao et al., 2009	Analysis of synonymous codon usage in classical swine fever virus	*Virus genes*	performed an Analysis of synonymous codon usage in classical swine fever virus (CSFV)
Fu., 2010	Codon usage bias in herpesvirus	*Archives of virology*	performed Codon usage bias in herpesvirus
Zhang et al., 2011	Analysis of codon usage and nucleotide composition bias in polioviruses	*Virology journal*	performed an Analysis of codon usage and nucleotide composition biasin polioviruses
Li et.al. 2018	Insights into the genetic and host adaptability of emerging porcine circovirus 3	*Virulence*	performed a deep codon usage analysis of porcine circovirus 3 (PCV3). PCV3 sequences were classified into two clades: PCV3a and PCV3b, confirmed by principal component analysis
Yao et al., 2019	Analysis of Synonymous	*BioMed research*	studied the codon usage pattern of *Flaviridae* viruses, causing

	Codon Usage Bias in Flaviviridae Virus	*international*	infection in humans and transmitted from mosquitos, and sandflies.
Li et al., 2019	Genetic analysis and evolutionary changes of the torque teno sus virus	*International journal of molecular sciences*	studied torque teno sus virus, the causative agent of porcine circovirus associated disease (PCVAD)
Baha et al., 2019	Comprehensive analysis of genetic and evolutionary features of the hepatitis E virus	*BMC genomics*	studied a Comprehensive analysis of genetic and evolutionary features of the hepatitis E virus
Biswas et al., 2019	Codon usage bias analysis of Citrus tristeza virus: Higher codon adaptation to citrus reticulata host	*Viruses*	performed Codon Usage Bias Analysis of Citrus tristeza virus
Tsai et al., 2019	Phylogeographic and genetic characterization of porcine circovirus type 2 in Taiwan from 2001–2017	*Scientific reports*	performed Phylogeographic and genetic characterization of porcine circovirus type 2 in Taiwan from 2001-2017
Krasovec and Filatop., 2019	Evolution of Codon Usage Bias in Diatoms		studied Codon Usage Bias in Diatoms belonging to 35 genera from 4 classes

CUB is thought to be caused by two key factors translational (selection) and mutational pressure (which is a bit of a misnomer) **(Bulmer., 1987)**. **Jenkins and Holmes., 2003** studied codon usage bias in a wide range of genetically and ecologically diverse human RNA viruses. **Coleman et.al., 2008** synthesized de novo large DNA molecules using hundreds of over-or underrepresented synonymous codon pairs to encode the poliovirus capsid protein. **Xu et al., 2008** performed an analysis of synonymous codon usage and the evolution of begomoviruses. **Tao et al., 2009**

performed an Analysis of synonymous codon usage in classical swine fever virus (CSFV). **Fu., 2010** performed Codon usage bias in herpesvirus. **Zhang et al., 2011** performed an Analysis of codon usage and nucleotide composition bias in polioviruses. **Liu et al., 2012** studied patterns and influencing factors of synonymous codon usage in Porcine Circovirus. **Belalov and Lukashev.,2013** analyze the relative impact of mutational pressure components on codon usage bias in RNA viruses. **Du et al., 2014** studied the selection on synonymous codons in mammalian rhodopsins and its possible role in optimizing translational processes. **Chen et.al., 2014** performed the analysis of synonymous codon usage in Porcine Circovirus to investigate the codon usage bias of PCV. **Roy et al., 2015** predicted codon and amino acid usage patterns in *Bifidobacterium* genus. **Kumar et al., 2016** examined the codon usage bias of Equine influenza viruses (EIVs) of H3N8 subtype, causing severe respiratory infections in horses, and are responsible for significant outbreaks worldwide. **Butt et al., 2016** examined codon usage bias in Zika virus by using various computational tools and software including CodonW, SPSS23, CAIcal server, eRCDI server, Recombination Detection Program server. **Dash et al., 2017** designed an epitope-based peptide vaccine against EBOV, using a combination of B-cell and T-cell epitope predictions, followed by molecular docking and molecular dynamics simulation approach. **Yao et al., 2019** studied the codon usage pattern of *Flaviridae* viruses, causing infection in humans and transmitted from mosquitos, and sandflies. **Khandia et al., 2019** analysed the codon usage and adaptation of Nipah virus and also the adaptation of viruses to their host. **Li et al., 2019** studied torque teno sus virus, the causative agent of porcine circovirus associated disease (PCVAD). **Baha et al., 2019** studied a Comprehensive analysis of genetic and evolutionary features of the hepatitis E virus. **Biswas et al., 2019** performed Codon Usage Bias Analysis of Citrus tristeza virus. **Tsai et al., 2019** performed Phylogeographic and genetic characterization of porcine circovirus type 2 in Taiwan from 2001-2017. **Krasovec and Filatop., 2019** studied Codon Usage Bias in Diatoms belonging to 35 genera from 4 classes.

CHAPTER 3
OBJECTIVES & SCOPE OF STUDY

The Canine Circovirus (CanineCV) has been detected in both symptomatic and asymptomatic dogs, and only vasculitis and histiocytic inflammation have been associated in some dogs. CanineCV has not been involved as the primary causative agent of the acute haemorrhagic diarrhoea syndrome in dogs, but might play an important role as a negative co-factor in the disease outcome in dogs with Canine parvo virus-2 infection. It is also suspected that younger individuals are more at risk of developing life-threatening symptoms than adult animals. Although there are potential vaccines available for these enteric viruses, there is currently no literature on a potential vaccine against the Circovirus for dogs. Moreover, the cure for CanineCV is currently unknown and treatment is non-specific. There is currently no literature on a potential vaccine against the Circovirus for dogs, but there is one in effect for porcine circovirus. Keeping in view all the above facts it is imperative to develop vaccine candidates that can prevent dogs from CanineCV infections by generating a robust immune response.

- Synonymous codon usage bias (SCUB) was done earlier on PCV, Pigeon CV & Duck CV. This will be the first attempt to evaluate codon usage pattern of CanineCV.
- No earlier attempts were made to find Mutational Pressure/Natural selection responsible for the host-specific adaptation of CanineCV.
- Currently, there is no vaccine against CanineCV.
- No earlier attempts were made for the development of vaccines using the codon usage bias manipulation strategy in CanineCV.

Keeping these challenges in view, the following research objectives were made: -

Research Objectives

1. Analysis of synonymous codon usage characteristics of CanineCV.

2. Evaluate factors such as mutational pressure and natural selection responsible for the host-specific adaptation of CanineCV.

3. Proteomics approach for prediction of vaccine candidates.

4. Codon usage bias manipulation strategy for the development of a vaccine. Generally, the degeneracy of the genetic code allows for 61 triplet codons that encode all 20 amino acids such that many of them are synonymous. Thus, codons encoding the same amino acid are termed synonymous codons. Interestingly, synonymous

codons are not randomly used in a given cellular species, and the preference for specific codons over other synonymous codons by various organisms or even in different gene groups of the same genome creates a bias in codon usage. This phenomenon is known as codon usage bias (CUB). Several factors, such as mutation pressure, natural selection, gene length, compositional constraints, replication, selective transcription, secondary protein structure, hydrophobicity, and gene function have been shown to influence codon usage patterns. CUB has been observed in a wide range of organisms, including viruses. Compared with prokaryotic and eukaryotic genomes, the small genome size, the interplay of codon usage between viruses and their hosts, and processes such as replication, protein synthesis, and transmission that depend on the host are expected to affect overall viral fitness and the avoidance of host cell responses during viral evolution. Synthetic attenuated virus engineering (SAVE), can be used to create non-virulent strains of viruses, which could provide a platform technology for vaccine development against a wide variety of viral diseases. The new vaccine strategy exploits the fact that the genetic code is degenerate in that most amino acids are encoded by more than one three-nucleotide codon (e.g., GCA, GCC, GCG, and GCT all code for the amino acid alanine). Many organisms, including viruses, tend to have biases toward certain synonymous codons (the “codon bias”) and codon pairs (the “codon pair bias”) in their genes. Codon pair bias refers to the preferential pairing of certain codons over others (e.g., in human genes the codon pair GCC-GAA encodes the adjacent amino acids alanine-glutamic acid less often than GCA-GAG, even though GCC and GAA are most used codons). Unfavourable host codon pairs introduced into a viral genome may adversely affect protein translation and thereby provide a method for producing attenuated viruses. In the present study, a novel epitope vaccine design is achieved by using a reverse vaccinology approach, which is going to help the medical world in many aspects. It will provide a better cure for CanineCV disease in dogs and specifically target CanineCV antigenic non-allergic epitopes. This investigation will provide economically (time as well as money) efficient strategy development for the elimination of CanineCV from dogs by using reverse vaccinology. This study opens more dimensions in the research domain for understanding the least studied CanineCV structure loopholes by molecular modelling of selected antigenic epitopes and their interaction with MHC class II alleles.

CHAPTER 4

METHODS & MATERIALS

Objective 1

1.1 Analysis of synonymous codon usage characteristics of CanineCV

1.1.1 The complete genomes of CanineCV strains

The complete genomes of CanineCV strains were retrieved in FASTA format from National Center for Biotechnology Information (NCBI). To minimize sampling bias in codon usage calculations, Coding Sequences (CDs) within the CanineCV genome were selected according to the following: (i) containing more than 300 bases (ii) starting with the start codon (iii) number of bases being a multiple of 3 and (iv) having no stop codon in the inner sequence. A multiple alignment of the Coding Sequences (CDs) was carried out using multiple alignment software BioEdit 7.2, Maga X MUSCLE (**MU**ltiple **S**equence **C**omparison by **L**og-**E**xpectation), or ClustalW. A phylogenetic tree was constructed using the neighbour-joining method in MEGA X software. The reliability of the neighbour-joining tree was calculated using 1,000 bootstrap replicates.

1.1.2 Nucleotide Composition Analysis

The following compositional properties were calculated for the coding sequences of the CanineCV genomes: (i) overall GC content (ii) overall frequency of nucleotides (A%, C%, U%, and G%) (iii) frequency of each nucleotide at the third site of the synonymous codons (A3S%, C3S%, U3S%, and G3S%) (iv) frequency of nucleotides G + C at the third synonymous codon positions (GC3S%) (v) frequency of nucleotides G + C at the third codon positions (GC3) and the mean of the frequency of both G + C at the first and second position (GC12). The codons AUG and UGG are the only codons for Methionine and Tryptophan, respectively, and the termination codons UAA, UAG, and UGA do not encode any amino acids. Therefore, these five codons were excluded from the analysis. Nucleotide composition was calculated using the program/software CodonW/EMBOSS/DAMBE/BioEdit/CAIcal Server.

1.1.3 Relative Synonymous Codon Usage

Relative Synonymous Codon Usage (RSCU) is a simple measure of non-uniform usage of synonymous codons in a coding sequence. RSCU values are the number of times a particular codon is observed, relative to the number of times that the codon

would be observed for a uniform synonymous codon usage (i.e. all the codons for a given amino-acid have the same probability). In the absence of any codon usage bias, the RSCU values would be 1.00. A codon that is used less frequently than expected will have an RSCU value of less than 1.00 and vice versa for a codon that is used more frequently than expected. The RSCU values of a synonymous codon refer to the relative probability of its observed frequency to its expected frequency, assuming that all codons for an amino acid are used equally. RSCU values were calculated using the software CodonW/CAIcal Server.

1.1.4 Effective number of codons (ENC)

The ENC value is used to detect the degree of codon usage bias. The ENC values range from 20 (only one synonymous codon is used, an extreme codon usage bias) to 61 (the synonymous codons are used equally, no bias). The effective number of codons (Nc) is a widely used index for characterizing codon usage bias because it does not require a set of reference genes as does codon adaptation index (CAI) and because of the freely available computational tools such as CodonW. The ENC value were calculated by CodonW.

1.1.5 Codon dinucleotide frequency analysis

The dinucleotide frequencies were calculated using the DAMBE software.

1.1.6 Gravy and aroma statistics

The Gravy value is the mean of the sum of the hydropathic indices of each amino acid which indicates the effect of protein hydrophobicity on codon usage bias. The Aroma value measures the effect of aromatic hydrocarbon proteins on codon usage bias. Gravy and aroma statistics were calculated by CodonW.

Objective 2

2.1 Evaluate factors such as mutational pressure and natural selection responsible for the host-specific adaptation of CanineCV

2.1.1 Enc-plot analysis

The ENC-GC3s plot is widely used to determine whether the codon usage of a gene is affected by mutation and selection. To study factors influencing codon usage bias. If the points fall below the expected curve, the codon usage is said to be affected by

selection pressure rather than mutation pressure, whereas mutational pressure is indicated when the data points fall onto the expected curve. ENC-plot analyses (ENC values against GC3s values), which consist of plotting GC3s values in the abscissa and the ENC values in the ordinate, are used to investigate the role of mutational pressure in codon usage bias. If the only factor driving the codon usage bias is mutation pressure, these points will be on the standard curve. Otherwise, other factors such as natural selection may play a crucial role the smaller the ENC value, the stronger the codon preference is. It is also accepted that ENC values <35 are indicative of genes with significant codon bias. ENC-plots (ENC value against GC3s value) was used to reveal the factors driving codon usage bias. If mutation pressure is the only factor, the point will lie on the standard curve.

2.1.2 Parity rule 2 (PR2) analysis

PR2 analysis is used to measure the effect of natural selection and mutation pressure. It is used to determine whether the biased codon selection was restricted to highly biased genes. There is no bias in the selection or mutation pressure when the plot lie on the center, where both coordinates are 0.5 plot analysis was performed to investigate the effects of mutation and natural selection on the codon usage of individual genes by exploring the relationship of the four-codon amino acid families, with A3/(A3+U3) plotted against G3/(G3+C3). The center of the plot is where A=U and G=C (PR2), indicating a balance between mutation pressure and natural selection.

2.1.3 Neutrality analysis

Neutrality analysis (GC12s against GC3s) was used to determine the dominant factor affecting codon usage bias. The values of GC12 and GC3 were calculated by the EMBOSS CUSP program/CodonW and then will be subjected to neutrality plot analysis.

2.1.4 Analysis of host-specific adaptation

(a) *Codon adaptation index (CAI) analysis*: The CAI values can estimate the degree of preference for codon usage of a gene based on the sequence of a known highly expressed gene. It represents the adaption of the virus to the host. The value of CAI value is between 0 and 1. A higher CAI value indicates

a stronger adaptation to the host. CAI is considered to be a determination of the codon usage tendency of the virus to its corresponding hosts and displays the expression level of the respective coding sequence. CAI values were calculated by the CAIcal Server. The reference database of the synonymous codon usage patterns was obtained from the Codon Usage Database (CUD).

(b) *Relative codon deoptimization index (RCDI) analysis:* RCDI is a measure of the tendency of the codon deoptimization of a virus to its hosts. The RCDI of the different genotypes of CanineCV was calculated by the RCDI/eRCDI Server. An RCDI value of 1 indicates that the virus is predominantly adapted to the host, while a value higher than 1 indicates less adaptability. The reference database will be the same used for CAI analysis.

2.1.5 *Correspondence Analysis (CoA)*

Correspondence analysis is an effective method of identifying the major trends in the codon usage patterns among virus coding sequences. Each coding region will be represented as a 59-dimensional vector corresponding to the RSCU value of each synonymous codon (excluding AUG, UGG, and stop codons). CoA was performed using CodonW.

2.1.6 Statistical Analysis

The correlation analysis was performed to identify the relationships between the GC, GC3s, ENC, the first two principal component axes, Aroma and Gravy, which were calculated using Spearman's rank correlation analysis, with an highly significant relationship of $p < 0.01$ and a significant relationship of $0.01 < p < 0.05$. Statistical/Correlation analysis was carried out to identify the factors influencing synonymous codon usage patterns by the statistical software SPSS22.0/SPSS26.0.

Table-1: Methodology used

Analysis	**Parameters to be measured**	**Software used**
Synonymous codon usage characteristics analysis of CanineCV	Codon usage bias analysis, *Composition of Nucleotide Analysis, RSCU, ENC analysis, Codon dinucleotide frequency analysis, Gravy and aroma statistics*	CodonW/EMBOSS/DAMBE/BioEdit/CAIcal Server CodonW/EMBOSS/DAMBE/BioEdit/CAIcal Server CodonW/CAIcal Server CodonW, DAMBE, CodonW
Evaluate factors such as mutational pressure and natural selection responsible for the host-specific adaptation of CanineCV	*Enc-plot analysis, Parity rule 2 (PR2) analysis, Neutrality analysis Codon adaptation index (CAI) analysis, Relative codon deoptimization index (RCDI) analysis, Correspondence Analysis (CoA), Statistical Analysis*	EMBOSSCUSP program/CodonW CAIcalServer (http://genomes.urv.cat/ CAIcal/RCDI/)RCDI/eRCDI Server (http://genomes.urv.cat/CAIcal/RCDI/), CodonW SPSS22.0/SPSS26.0

Objective 3

3.1. In-silico prediction of multi-epitope vaccine construct

3.1.1 Proteomic data retrieval

ViPR database was used as it is based on IEDB for database assessment that assisted us in gathering information regarding viral proteome. The viral replicase and capsid proteins structural sequences of Canine circovirus were retrieved from the NCBI-Proteomic database with accession ID's: QFU80922.1 replicase [*Canine circovirus*] and QBQ20241.1 capsid Protein [*Canine circovirus*] (Table 5).

3.1.2 Epitopes selection and structure prediction

The NetMHCpan-4.1 server predicts peptides/ epitopes from viral proteomic determinants by using artificial neural networks to varied MHC molecules with familiar sequences (ANNs) **(Reynisson et al., 2020)**. To predict the epitopes dog MHC molecules: DLA-8803401, DLA-8850101, and DLA-8850801 were selected. The predicted epitopes were selected based on Bind Level i.e. if they are strong binders or weak binders. The weak and strong binders were determined by using the

default parameters of the webserver. By default, a peptide was predicted as a strong binder if its percentage rank is below 0.5% and a peptide was predicted as a weak binder if its percentage rank is above 0.5% but below 2% rank. VaxiJen was also used in this study to determine the antigenic characterization of the selected epitope. Then selected epitope structural prediction was conducted by using PEP-FOLD, a De novo approach aimed at predicting peptide structures from amino acid sequences. This approach couples the deep learning algorithm and a coarse-grained force field based on amino acids to determine the conformations of consecutive amino acid residues as per their physiochemical relationships in secondary and tertiary folding **(Lamiable et al., 2016)**.

3.1.3 Molecular docking and simulation at epitope-DLA level

For molecular docking between DLA proteins and epitopes, Patchdock **(Schneidman-Duhovny et al., 2005)** and Autodock Vina **(Trott and Olson.,2009)** were used. The structures for DLA allelic sets were retrieved from the RCSB-PDB database. DLA alleles DLA-8803401 and DLA-8850801 were considered for performing molecular docking and simulation analysis. The PDB ID of DLA-8803401 and DLA-8850801 are '7CJQ' and '5F1I' respectively. For molecular docking and simulation studies, the PDB file of the epitopes and DLA were used. PatchDock assists the user to determine atomic contact energies and Autodock vina assists the user to determine binding energies for perfectly docked complexes. While for stability analysis of docked complexes molecular simulation was performed by deploying Gromacs **(Abraham et al., 2015)**.

3.1.4 Full-fledged *in-silico* vaccine construction

Codon adaptation was analyzed and *in silico* cloning was conducted for selected epitopes interacting with DLA allelic determinants after joining them with the proper number of required adjuvants. The J-Cat server **(Grote et al., 2005)** was used for this purpose. The main full-fledged vaccine structure was then subjected to stability analysis and physio-chemical characteristics determination by deploying the 'ExpasyProtParam' and Ramachandran plot analysis tool (MolProbity server). The solubility of the vaccine was evaluated by the SOLpro webserver **(Magnan et al., 2009)**. The SOLpro web server predicts the solubility of proteins with 74% accuracy after expression in *Escherichia coli* **(Magnan et al., 2009)**.

3.1.5 Vaccine construct prediction and affirmation

The secondary structure of the vaccine was determined by using the PSIPRED webserver **(Buchan and Jones, 2019)**. This web server helps in the prediction of beta-sheets, alpha helices, and coils in proteins by using feed-forward neural networks **(Buchan and Jones, 2019)**. The tertiary structure of the vaccine was determined by using the I-TASSER server which uses iterative template-based fragments assembly simulations and multiple threading approaches to predict the 3D model of proteins **(Yang et al., 2014)**, uses iterative template-based fragments assembly simulations and multiple threading approaches to predict the 3D model of proteins. Finally, the tertiary structure of the modeled vaccine was validated by the MolProbity webserver **(Chen et al., 2010)**.

3.1.6 Codon Adaptation and vaccine *in-silico* cloning

For efficient cloning and expression of the vaccine in expression vectors, codon optimization is imperative. JCAT webserver was used for codon optimization of vaccine for expression in *E. coli*K12strain **(Grote et al., 2005)**. The 'Snap Gene' restriction cloning module was used for *in-silico* cloning of vaccine.

3.2. In-silico Epitope based vaccine prediction and wet-lab validations

3.2.1 Protein Sequence Retrieval

The sequence of the CanineCV proteins., capsid and replicase were obtained from the NCBI GenBank with accession numbers QBQ20241.1 and QFU80922.1, respectively, in FASTA format. Further, the protein sequences of the CanineCV capsid and replicase were obtained from China, USA, and Germany: CanineCVstrain XF16 (China), CanineCV strain OH19098-1 (USA), and CanineCV strain FUBerlin-JRS (Germany). The accession number for replicase and capsid are AVT56110.1, AWN93250.1, and ALG63389.1, and AVT56111.1, AWN93251.1, and ALG63390.1, respectively. The multiple sequence alignment for the four strains corresponding to capsid and replicase proteins indicates the high sequence identity (i.e., >94%) among the chosen strains from different regions, see Supplementary Information Table S1 and S2.

3.2.2 Antigenicity, Toxicity, Conservancy, and IFN-γ Analysis

The epitope screening analysis was carried out by NetMHCIIPan 4.0 server **(Reynisson et al., 2020)** which allows epitope prediction based on a neural network approach. Subjecting the protein sequences to the NetMHCIIPan server led to the identification of a total of 289 epitopes from replicase and 256 epitopes from capsid for each strain. Further, conservancy and antigenicity analyses were conducted. To validate the results of predicted epitopes obtained by the NetMHCIIPan server, an independent approach employing the Genscript Optimum Antigen design tool was also used. This tool employs a high-end algorithm to search various protein databases to design epitopes with desired antigenicity and specificity. The algorithm considers various parameters such as sequence length, secondary structure regions, surface orientation, hydrophobicity/hydrophilicity, and avoidance of sequence motifs such as GTP binding sites, RGD motif, and SH2 domains while predicting the antigenic epitopes. An extra cysteine is added to N- terminus or C-terminus for the identified epitopes to facilitate conjugation while predicting the antigens using the Optimum Antigen design tool. For conservancy analysis of the predicted epitopes, IEDB Epitope Conservancy Tool (http://tools.iedb.org/conservancy/) was used for CanineCV strain XF16 (China), CanineCV strain OH19098-1 (USA), and CanineCV strain FUBerlin-JRS (Germany). For these chosen strains of CanineCV, the accession numbers are AVT56110.1, AWN93250.1, and ALG63389.1 for replicase, and are AVT56111.1, AWN93251.1, and ALG63390.1 for capsid, respectively. Finally, the IFN epitope server was used for interferon-γ inducing ability analysis of the predicted epitopes **(Dhanda et al., 2013)**.

3.2.3 Vaccine Engineering and Determination of its Physiochemical Properties

After the selection of suitable epitopes from the capsid and replicase proteins of CanineCV, the final multi-epitope vaccine construct was designed. GPGPG sequence was used as a linker to join RS09 and flagellin adjuvants with the selected epitopes. Pan DR epitope (PADRE) sequence was also added to improve the immunogenicity as well as the stability of the CanineCV vaccine construct. The ExPASy ProtParam webserver was used to predict the physicochemical properties such as stability, isoelectric point, molecular weight, aliphatic index as well as several other properties of the MEV construct **(Gasteiger et al., 2005)**. The antigenicity of the predicted MEV

construct was also analyzed by the Vaxijen webserver **(Doytchinova and Flower., 2007)**. This server predicts the antigenicity of the viral, fungal, tumor, or bacterial peptide sequences based on their physicochemical properties. Further, the ToxinPred server was used for toxicity analysis **(Gupta et al., 2013)**. The ToxiPred web server can predict the toxicity of peptides/proteins and allow mutations to modulate the toxicity of the peptides.

3.2.4 Molecular Modeling, Docking, and Molecular Dynamic Simulations of Designed Multi Epitope Vaccine with TLR-5

To predict the 3-D structures of MEV and Canine TLR-5, the I-Tasser program was used **(David et al.,2022; Jumper et al., 2021)**. The tertiary structures of the MEV and TLR-5 were further validated by generating a Ramachandran plot using the ProCheck webserver. For docking of the MEV and TLR-5 complex, the HADDOCK server **(van Zundert et al., 2016)** was used. Further, all the Molecular Dynamic (MD) simulations were performed with the GROMACS2019 simulation program **(Abraham et al., 2015)** TLR-5 complexed with MEV was placed in a cubic box and solvated with TIP3P water molecules creating a solvent layer at least 12 Å thick. The Amber ff99SB-ILDN **(Lindorff-Larsen et al., 2010)** force field was used to model the parameters of proteins. The charge was neutralized adding an appropriate number of K^+ ions. The extra K^+ Cl^- ions were added to achieve the bulk ionic strength of 0.15 M using the Joung-Cheatham ion model **(Joung et al., 2008)**. The simulation box contains 453556 water molecules, 1216 K^+ ions, and 1229 CL^- ions. The total number of atoms in the system was 1377736. The system was first minimized with 50000 steps of the steepest descent method with 1000 kJ/mol nm^2 position restraint on protein-heavy atoms. Further, minimization was carried out without any restraint on protein. Equilibration of each system was carried out in a phased manner. Firstly, 100ps NVT simulation was carried out with restraint on heavy atoms of the protein. In the second step, 100psNPT simulation with restraint on heavy atoms of the protein was performed. Production simulations were run using the NPT ensemble for 100ns. The temperature was maintained at 300K using velocity rescaling with a coupling time of 0.1ps. The pressure was maintained at 1atm for NPT simulations using a Parrinello–Rahmanbarostat **(Parrinello and Rahman., 1981)** with a coupling time of 2ps. Equations of motion were integrated using the leapfrog algorithm with a time step of 2.0fs. The total electrostatic interactions were evaluated using the particle

mesh Ewald (PME summation). Coulomb and Van der Waals cut-offs of 1.0nm were employed. Periodic boundary conditions in all directions were employed to mimic the bulk behavior. Bond lengths with hydrogen were constrained with the LINCS algorithm **(Darden et al., 1993, Hess et al., 1997)**. Coordinates were kept collected in trajectory files every 10ps. Trajectory processing and most of the analysis were performed using the GROMACS tools. The visualization and molecular graphics images were created using the PYMOL **(Humphrey et al., 1996)** and VMD software. Graphs were plotted using in-house python scripts. Our group has successfully employed a similar modeling and MD simulation protocol for studying the structural stability of proteins **(Oliva et al., 2021, Grewal et al., 2021)** and the Nucleic acids system **(Chawla et al., 2021, Kalra et al., 2021)**.

3.2.5 Peptide Synthesis and Conjugation with Carrier Protein

One of the potential antigenic peptides with RVRRHARASRRSYRC sequence was synthesized using GenScript's microwave-based PepPower™ technology merging Solid-phase peptide synthesis, Liquid-phase peptide synthesis, and Microwave technology. The peptide concentration was measured using NanoDrop Spectrophotometer at 280nm. The purity was assessed with SDS-PAGE and the antigen with a purity of >70% was selected for further analysis. The cysteine residue was added at the N-terminal of the purified peptide and was subsequently conjugated with Keyhole Limpet Hemocyanin (KLH) using the m-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) method.

MATERIALS AND PROTOCOLS

Antigen details for Canine Circovirus

Virus	Protein	Identifier	Epitope
Canine circovirus	Capsid	QBQ20241.1 capsid Protein	**RVRRHARASRRSYRC**

No	**Antigenic Determinant**	**Length**	**Antigenicity/ Surface/Hydrophilicity**	**Disordered Score**	**Synthesis**	**Oryctolagus_cuniculus blast**
1	RVRRHARASRRSYRC	15	2.11/1.00/1.04	0.3643	Easy	46%

ELISA titer >1:128,000 with immunogen and ≥2mg purified antibody*2
(Purity>80%) can be guaranteed.

Procedure:

1. For toxicity analysis: in the mammalian cell line.
2. For Antibody development: This experiment was performed in rabbits. ELISA was also performed to determine the titer of the antibodies produced in rabbits against your peptides. Also, determine the antigenicity of the peptides using algorithms.

3. Animal immunization

4. Antibody purification

5. Custom Polyclonal Antibody Development Against Peptide Antigen

Sequence:

RVRRHARASRRSYRC

The blue regions are predicted as antigenic determinants.

The underline regions are predicted as disordered regions.

The regions which background gray are predicted as transmembrane regions.

Antigen Designed:

No	Start	Antigenic Determinant	Length	Antigenicity/Surface /Hydrophilicity	Disordered Score	Synthesis	Oryctolagus_ cuniculus blast
1	1	RVRRHAR ASRRSYRC	15	2.11/1.00/1.04	0.3643	Easy	46%

1. An extra "C" (high-lighted as green) is added to the C-terminus (or N-terminus) to facilitate conjugation.

2. Positively charged residues (K,R,H) are in blue. Negative charged residues (D,E) are in red.

3. For a basic peptide, initially try to dissolve the peptide in water; if the peptide does not dissolve, try 10% and higher solutions of acetic acid; if the peptide still does not dissolve, add TFA(<50ul) to solubilize the peptide and dilute to 1ml with deionized water.

For an acidic peptide, initially try to dissolve the peptide in water, if the peptide does not dissolve, add NH_4OH(<50ul) to solubilize the peptide and dilute to 1ml with deionized water.

4. Disorder peptide note: The linear conformation of a disordered peptide is more

similar to its natural conformation in the folded protein.

ELISA protocol

Reagents

Coating Buffer (1XPBS Buffer)

8.5 g NaCl

1.4 g Na_2HPO_4

0.2 g NaH_2PO_4

1000 ml dd H_20

Adjust pH to pH 7.4

Store at 4°C

Washing Buffer

0.5 ml Tween 20

1000 ml PBS Buffer

Store at 4°C

Blocking Buffer

100 ml Washing Buffer

1 g BSA

Store at 4°C

Stop Buffer

8.3 ml 12 mol/L HCl

91.7 ml ddH_2O

Store at 4°C

TMB Reagent (GenScript Cat.No M00078)

Indirect ELISA protocol

Procedure

Coating

1. Dilute the antigen with *Coating Buffer* and coat appropriate wells of ELISA plate with the antigen by pipeting 100 µl of the diluted solution. The concentration of coated antigen ranges from 1 -10 µg/ml.

2. Cover the plate with an adhesive plastic and incubate for 2 hours at 37^oC or

4°C overnight.

3. Remove the antigen coating solution from the wells of plate by flicking the plate over a sink.

4. Wash the plate three times by adding the wells with 200 µl of *Washing Buffer*.

Blocking

5. Add 200 µl of *Blocking Buffer* to block the non-specific binding sites in the coated wells.

6. Cover the plate with an adhesive plastic and incubate for at 2 hours at 37°C or overnight.

7. Remove the *Blocking Buffer* from the wells of plate by flicking the plate over a sink.

8. Wash the plate three times by adding the wells with 200 µl of *Washing Buffer.*

Incubation

Dilute the primary antibody or antiserum with *Blocking Buffer* and add 100 µl of the diluted antibody to each well of the plate. The concentration of primary antibody is based on the manufacturer's instructions

9. Cover the plate with an adhesive plastic and incubate for 1 hour at 37°C or overnight at 4°C.

10. Remove the diluted primary antibody solution from the wells of plate by flicking the plate over a sink.

11. Wash the plate three times by adding the wells with 200 µl of *Washing buffer*.

12. Dilute the HRP-conjugated secondary antibody with *Blocking Buffer* and add 100 µl of the diluted secondary antibody to each well of the plate.

13. Cover the plate with an adhesive plastic and incubate for 30 minutes at 37°C.

14. Remove the diluted secondary antibody from the wells of plate by flicking the plate over a sink.

15. Wash the plate five times by adding the wells with 200 µl of *Washing Buffer.*

Detection

16. Add 100 µl of the *TMB Reagent* per well with a multichannel pipete or a multipipete.

17. After sufficient color development, add 100 µl of *Stop Buffer* to the wells.Note: 15-30 minutes is enough for color development.

18. Read the absorbance of each well with a plate reader.

3.2.6 *In-vivo* Validation for Antibody Development and Assessment

The KLH-conjugated peptide was inoculated in the New Zealand rabbit (n=2) following GeneScript's PolyExpress immunization method. This method was used in anti-KRV envelope polyclonal antibody generation **(Tree et al., 2016)**. The antibodies were retrieved and the purity of the polyclonal antibody was determined by performing SDS-PAGE. Further, the concentration of the antibody was measured by NanoDrop Spectrophotometer at 280nm. Afterward, the antibody titer was determined with the indirect ELISA using IgG as the control. The free peptide was used as the antigen at the coating concentration of 4ug/ml, and 100µl/well for ELISA development. The anti-rabbit horseradish peroxidase (HRP) coupled IgG was used as the secondary antibody. The OD_{450nm} value was measured using an ELISA plate reader.

Objective 4

4.1 Codon usage bias manipulation strategy for the development of a vaccine

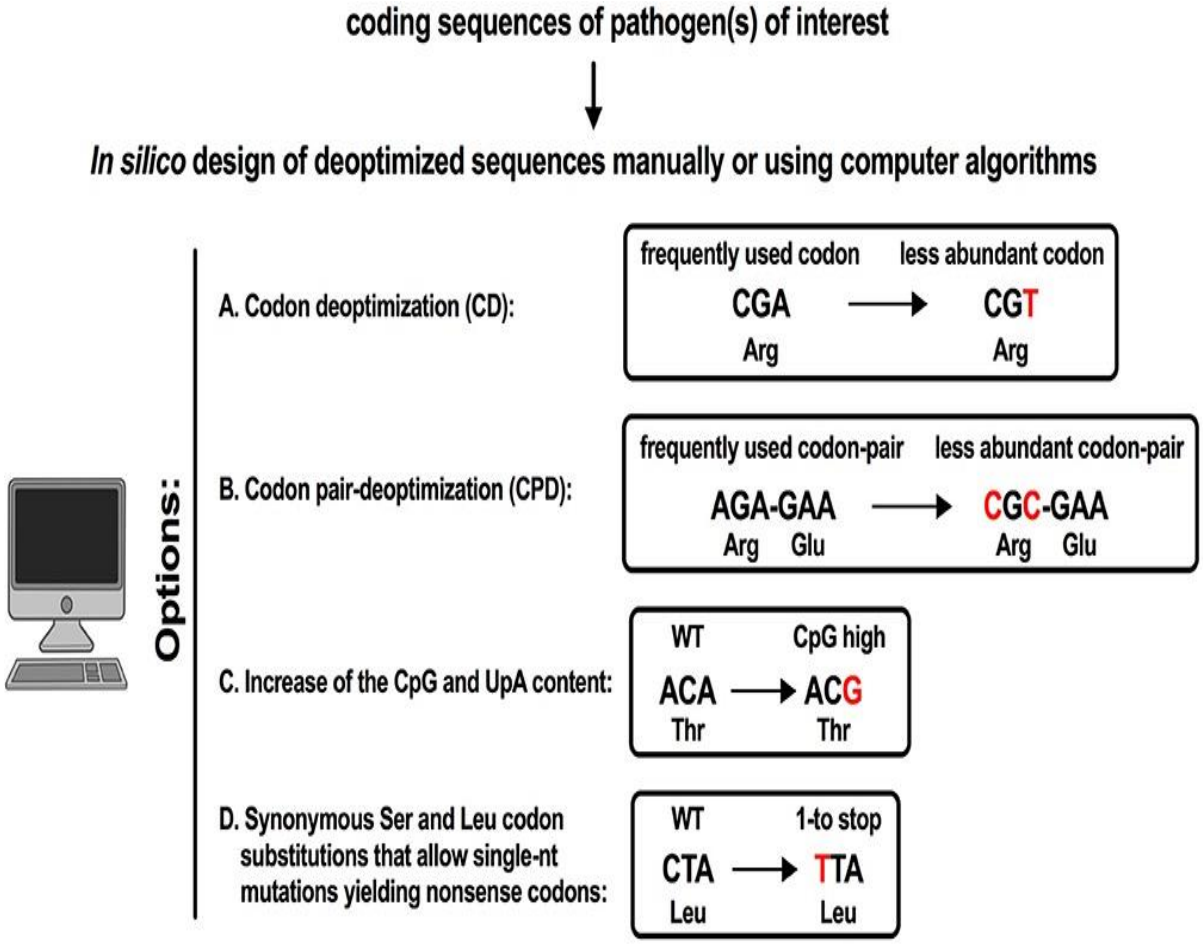
4.1.1 Four strategies are used to generate genome-scale deoptimized viruses

(a) codon deoptimization (CD)

(b) codon-pair deoptimization (CPD)

(c) increase of the CpG and UpA content

(d) synonymous Serine and Leucine codon substitutions that allow single-nt mutations yielding non-sense codons. The synonymous mutations generated by the deoptimization process are indicated in red.



CHAPTER 5
RESULTS & DISCUSSION

Table 2. Canine Circovirus variants sequences were used for this study

Ser No	Accession Number	Country	Year of Isolation	Strain	Virus Specie	Host Specie	Gene/CDS
01	JQ821392	USA	2012	Strain 214	Canine Circovirus	Dog	Cap/Rep
02	KC241982	USA	2012	UCD1-1698	Canine Circovirus	Dog	Cap/Rep
03	KC241983	USA	2012	UCD3-478	Canine Circovirus	Dog	Cap/Rep
04	KC241984	USA	2012	UCD2-32162	Canine Circovirus	Dog	Cap/Rep
05	NC020904	USA	2013	UCD1-1698	Canine Circovirus	Dog	Cap/Rep
06	MF457592	USA	2017	OH19098-1	Canine Circovirus	Dog	Cap/Rep
07	MK033608	South America	2018	UBA-Baires	Canine Circovirus	Dog	Cap/Rep
08	KJ530972	Italy	2014	Bari/411-13	Canine Circovirus	Dog	Cap/Rep
09	KT734812	Italy	2015	CB6293/1-14	Canine Circovirus	Dog	Cap/Rep
10	KT734813	Italy	2015	AZ2972-13	Canine Circovirus	Dog	Cap/Rep
11	KT734814	Italy	2015	TE4016-13	Canine Circovirus	Dog	Cap/Rep
12	KT734815	Italy	2015	AZ4133/1-13	Canine Circovirus	Dog	Cap/Rep
13	KT734816	Italy	2015	AZ4438-13	Canine Circovirus	Dog	Cap/Rep
14	KT734817	Italy	2015	AZ5212/1-14	Canine Circovirus	Dog	Cap/Rep
15	KT734818	Italy	2015	AZ5212/2-14	Canine Circovirus	Dog	Cap/Rep
16	KT734819	Italy	2015	AZ5586-13	Canine Circovirus	Dog	Cap/Rep
17	KT734820	Italy	2015	AZ663/1-13	Canine Circovirus	Dog	Cap/Rep
18	KT734821	Italy	2015	TE6685/1-13	Canine Circovirus	Dog	Cap/Rep
19	KT734822	Italy	2015	TE7482-13	Canine Circovirus	Dog	Cap/Rep
20	KT734823	Italy	2015	PE8575/1-13	Canine Circovirus	Dog	Cap/Rep
21	KT734824	Italy	2015	AZ663/2-13	Canine Circovirus	Dog	Cap/Rep
22	KT734825	Italy	2015	TE6685/2-	Canine	Dog	Cap/Rep

				13	Circovirus		
23	KT734826	Italy	2015	PE8575/2-13	Canine Circovirus	Dog	Cap/Rep
24	KT734827	Italy	2015	AZ4133/2-13	Canine Circovirus	Dog	Cap/Rep
25	KT734828	Italy	2015	CB6293/2-14	Canine Circovirus	Dog	Cap/Rep
26	MH454599	Itlay	2018	09-10F/2011	Fox Circovirus	Red Fox	Cap/Rep
27	KY388483	China	2016	strain WM60	Canine Circovirus	Dog	Cap/Rep
28	KT946839	China	2015	JZ98/2014	Canine Circovirus	Dog	Cap/Rep
29	KY388484	China	2016	strain WM48	Canine Circovirus	Dog	Cap/Rep
30	KY388485	China	2016	strain WM46	Canine Circovirus	Dog	Cap/Rep
31	KY388487	China	2016	strain LA237	Canine Circovirus	Dog	Cap/Rep
32	KY388491	China	2016	strain WM63	Canine Circovirus	Dog	Cap/Rep
33	KY388492	China	2016	strain WM66	Canine Circovirus	Dog	Cap/Rep
34	KY388493	China	2016	strain WM62	Canine Circovirus	Dog	Cap/Rep
35	KY388497	China	2016	strain WM84	Canine Circovirus	Dog	Cap/Rep
36	KY388499	China	2016	strain WM79	Canine Circovirus	Dog	Cap/Rep
37	KY388500	China	2016	strain WM77	Canine Circovirus	Dog	Cap/Rep
38	KY388501	China	2016	strain WM76	Canine Circovirus	Dog	Cap/Rep
39	KY388502	China	2016	strain WM74	Canine Circovirus	Dog	Cap/Rep
40	KY388503	China	2016	strain WM72	Canine Circovirus	Dog	Cap/Rep
41	KY388480	China	2016	strain GL33	Canine Circovirus	Dog	Cap/Rep
42	KY388481	China	2016	strain JZ50	Canine Circovirus	Dog	Cap/Rep
43	KY388482	China	2016	strain GL51	Canine Circovirus	Dog	Cap/Rep
44	KY388486	China	2016	strain	Canine	Dog	Cap/Rep

				LA280	Circovirus		
45	KY388488	China	2016	strain LA128	Canine Circovirus	Dog	Cap/Rep
46	KY388489	China	2016	strain JZ85	Canine Circovirus	Dog	Cap/Rep
47	KY388490	China	2016	strain JZ82	Canine Circovirus	Dog	Cap/Rep
48	KY388494	China	2016	strain YL11	Canine Circovirus	Dog	Cap/Rep
49	KY388495	China	2016	strain XXT243	Canine Circovirus	Dog	Cap/Rep
50	KY388496	China	2016	strain XXT242	Canine Circovirus	Dog	Cap/Rep
51	KY388498	China	2016	strain WM83	Canine Circovirus	Dog	Cap/Rep
52	MG279132	China	2017	strain 398	Canine Circovirus	Dog	Cap/Rep
53	MG279133	China	2017	strain 395	Canine Circovirus	Dog	Cap/Rep
54	MG279134	China	2017	strain 394	Canine Circovirus	Dog	Cap/Rep
55	MG279136	China	2017	strain 391	Canine Circovirus	Dog	Cap/Rep
56	MG279137	China	2017	strain 390	Canine Circovirus	Dog	Cap/Rep
57	MG279138	China	2017	strain 388	Canine Circovirus	Dog	Cap/Rep
58	MG279139	China	2017	strain 384	Canine Circovirus	Dog	Cap/Rep
59	MG266899	China	2017	CD17/2016	Canine Circovirus	Dog	Cap/Rep
60	MF797786	China	2017	strain XF16	Canine Circovirus	Dog	Cap/Rep
61	MG279118	China	2017	strain 102	Canine Circovirus	Dog	Cap/Rep
62	MG279119	China	2017	strain 199	Canine Circovirus	Dog	Cap/Rep
63	MG279120	China	2017	strain 198	Canine Circovirus	Dog	Cap/Rep
64	MG279121	China	2017	strain 186	Canine Circovirus	Dog	Cap/Rep
65	MG279122	China	2017	strain 176	Canine Circovirus	Dog	Cap/Rep
66	MG279123	China	2017	strain 185	Canine	Dog	Cap/Rep

					Circovirus		
67	MG279125	China	2017	strain 182	Canine Circovirus	Dog	Cap/Rep
68	MG279126	China	2017	strain 181	Canine Circovirus	Dog	Cap/Rep
69	MG279127	China	2017	strain 180	Canine Circovirus	Dog	Cap/Rep
70	MG279128	China	2017	strain 179	Canine Circovirus	Dog	Cap/Rep
71	MG279129	China	2017	strain 178	Canine Circovirus	Dog	Cap/Rep
72	MG279130	China	2017	strain 177	Canine Circovirus	Dog	Cap/Rep
73	MG279131	China	2017	strain 202	Canine Circovirus	Dog	Cap/Rep
74	MG279135	China	2017	strain 201	Canine Circovirus	Dog	Cap/Rep
75	MG279140	China	2017	strain 205	Canine Circovirus	Dog	Cap/Rep
76	MG279141	China	2017	strain 204	Canine Circovirus	Dog	Cap/Rep
77	MK731981	China	2019	strain C24	Canine Circovirus	Dog	Cap/Rep
78	MK731982	China	2019	strain K1	Canine Circovirus	Dog	Cap/Rep
79	MN128702	China	2019	strain NC21	Canine Circovirus	Dog	Cap/Rep
80	MK944079	China	2019	strain C79	Canine Circovirus	Dog	Cap/Rep
81	MK944080	China	2019	strain C85	Canine Circovirus	Dog	Cap/Rep
82	KP260925	Netherlands	2014	VS7100001	Canine Circovirus	Dog	Cap/Rep
83	KP260926	Netherlands	2014	VS7100003	Canine Circovirus	Dog	Cap/Rep
84	KP260927	Netherlands	2014	VS7100005	Canine Circovirus	Dog	Cap/Rep
85	KF887949	Germany	2013	strain Ha13	Canine Circovirus	Dog	Cap/Rep
86	KT283604	Germany	2015	FUBerlin-JRS	Canine Circovirus	Dog	Cap/Rep
87	KP941114	Croatia	2015	strain 55590	Fox Circovirus	Fox	Cap/Rep
88	MK424788	Brazil	2020	strain D1056	Canine Circovirus	Dog	Cap/Rep

Table 3: Nucleotides composition of CanineCV genome at third codon position (T3%, G3%, C3% and A3%), overall AT% and GC% composition along with GC12 (the average of GC1 and GC2) and GC3% analysis. *(Supplementary Table)

Table 4: RSCU values

S.No.	Amino acid	Codons	Coding sequences (CDSs)	
			C	R
1	Phenylalanine (F)	TTT	0.6345	1.25
		TTC	1.3654	0.7494
2	Leucine (L)	TTA	0.4469	0.5103
		TTG	0.9232	1.6306
		CTT	0.5977	1.1133
		CTC	1.0821	1.0514
		CTA	1.0748	0.0706
		CTG	1.8747	1.6237
3	Isoleucine (I)	ATT	0.7359	1.215
		ATC	0.8424	1.175
		ATA	1.4214	0.6088
4	Valine (V)	GTT	0.8074	1.2879
		GTC	0.8324	1.0201
		GTA	1.2438	0.3155
		GTG	1.1158	1.3767
5	Serine (S)	TCT	1.3975	1.7173
		TCC	0.6125	0.1988
		TCA	2.0131	0.5917
		TCG	0.4743	0.8141
		AGT	0.7614	0.9998
		AGC	0.7411	1.6778
6	Proline (P)	CCT	0.7839	1.5291
		CCC	0.1906	1.1443
		CCA	2.2456	0.1546
		CCG	0.7801	1.1719
7	Threonine (T)	ACT	1.0025	0.8869
		ACC	0.5531	1.6699
		ACA	2.0829	0.2641
		ACG	0.3614	1.1789
8	Alanine (A)	GCT	1.1787	1.598
		GCC	1.0211	1.0846
		GCA	1.5243	0.3621
		GCG	0.2757	0.9554
9	Tyrosine (Y)	TAT	1.0886	1.0205
		TAC	0.9113	0.9794
10	Histidine (H)	CAT	1.4577	1.2806
		CAC	0.5422	0.7193
11	Glutamine (Q)	CAA	1.4125	0.725
		CAG	0.5875	1.275
12	Asparagine (N)	AAT	1.0714	0.9169
		AAC	0.9285	1.083

13	Lysine (K)	AAA	1.6914	0.816
		AAG	0.3085	1.184
14	Aspartic acid (D)	GAT	0.8274	1.2857
		GAC	1.1725	0.7142
15	Glutamic acid (E)	GAA	1.1948	0.4393
		GAG	0.8051	1.5606
16	Cysteine (C)	TGT	0.0227	0.8746
		TGC	0.0454	1.1253
17	Arginine (R)	CGT	0.2766	0.8916
		CGC	1.0568	0.9323
		CGA	0.2839	1.2023
		CGG	0.6218	2.2048
		AGA	2.3553	0.6986
		AGG	1.4054	0.07008
18	Glycine (G)	GGT	0.5603	1.2742
		GGC	0.7602	1.091
		GGA	1.99	0.62
		GGG	0.6891	1.0145
			0.967205085	0.999938644

Codon usage bias of *cap* & *rep* gene and Nucleotide composition of 88 CaCV strains were compared. The overall nucleotide composition of CDSs at the 3rd codon position (A3% C3% G3% T3%) along with overall GC% and AT%, GC12, and GC3 were analyzed. The Aroma and Gravy values, effective no of codons (ENC), and relative synonymous codon usage (RSCU) values were also analysed using codon W 1.4.2 and the CAIcal server. The results show that CaCV genome is AT rich and A/T ending codons were preferred then GC ending codons. Nc-GC3 plot of CanineCV reveals that selection pressure was dominant over mutational pressure. The correlation analysis between CAI, Gravy, and Aroma indicates natural selection over mutational pressure. The RSCU values of *Cap* & *Rep* genes were analysed to find out overrepresented codons. We identified in this study the Codon usage analysis and selection pressure for the evolutionary perspective. Nucleotide composition of CDSs was determined at the 3rd codon position (A3% C3% G3% T3%) along with overall GC% and AT%, GC12, and GC3. The Aroma and Gravy values, ENC values, and RSCU values were also analysed using codon W 1.4.2 and the CAIcal server.

Nucleotide composition and codon usage bias of *cap* & *rep* gene of 88 CaCV strains were compared. The overall nucleotide composition of CDSs at the third codon

position (A3% C3% G3% T3%) along with overall GC% and AT%, GC12, and GC3 were analysed. The Aroma and Gravy values, effective no of codons (ENC), and relative synonymous codon usage (RSCU) values were also analysed using codon W 1.4.2 and CAIcal server.

The results show that CaCV genome is AT rich and A/T ending codons were preferred then GC ending codons. Nc-GC3 plot of CanineCV reveals that selection pressure was dominant over mutational pressure. The correlation analysis between CAI, Gravy, and Aroma indicates natural selection over mutational pressure. The RSCU values of *Cap* & *Rep* genes were analysed to find out overrepresented codons.

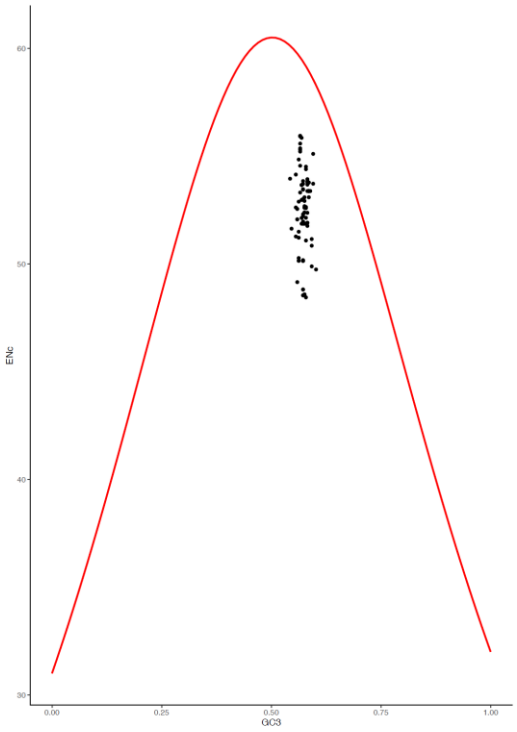


Fig 1.: Nc-GC3 plot of CanineCV reveals that selection pressure was dominant over mutational pressure

5.1. In-silico Prediction of multi-epitope vaccine construct

5.1.1. Proteome structure and sequence

We retrieved FASTA sequences of protein from the NCBI database and summarized them in Table 5.

Table 5: Retrieved FASTA sequences

>QFU80922.1 replicase [*Canine circovirus*]
MAQAQVDQRGRDSRRGNPVRRWCFTINNPTPEEEEAVKNLAPDAKYLICGREVGENGTPHL
QGFVNLKKTTRMGALKARLGGRGHFEPARGDDCSNKDYCSKGGDILIESGEVSRQGKRNDL
HDAVEKLRETKSLAAVAAAYPETYVKFSRGLRELLLISPEMTTPRNWKTEVEVLCGPPGCGK
SRYCMETAPDAYWKPRGKWWDGYDGHQDVILDDFYGWLPFDDMLRLCDRYPLRVETKGG
TMNFVARRVFITSNRLPHEWYSDEIGNKDALYRRLTLIKVWDGGNFIPVPHFMFPHMYNY

>QBQ20241.1 capsid Protein [*Canine circovirus*]
MRVRRHARASRRSYRTRPLNRYRRRRQNRFKLFHLRLRRTLTADWPTAPVKPTNDPQTETPL
LWNFDHLSFKLTDFLQTSHGTGDYQHLPPFRFFKFKKVYIRAKWINWPRTLMENVLGRTALD
LDGEDQGRGNAQRSHLDPGCVPGRLEPPKDPNKAPFIYDPLQDRSSSRSFNMASGFKRGLTPK
PMFTQDITSPSATAPWLTRGTPWVSVIQGANMVWNGLSISLRQMKDMRPTTPDTSTSQIPQVQ
YDISAYIAFKEFDYETGRQL

5.1.2. Predicted epitopes and DLA structure

NetMHCpan 4.1 server was deployed to find out interacting epitopes, lowest values were preferred while selecting epitopes based on prediction scores. For viral replicase 294 epitopes interacting with each DLA, allele was identified, while for viral capsid protein 261 epitopes interacting with each DLA allele were identified. Out of these 555 epitopes, the best 20 were considered based on epitope prediction scores. VaxiJen server was also used for determining antigenicity with a threshold value of 0.7. Five epitopes were selected out of 20 epitopes based on their antigenicity (Table 6.). Structures were constructed by deploying the PEP-FOLD server, which follows de-novo criteria for predicting perfect conformation for epitopes. The structures for DLA allelic sets were retrieved from the RCSB-PDB database, two major DLA alleles were considered during the investigation DLA-8803401 and DLA-8850801 with PDB IDs '7CJQ' and '5F1I' respectively. Both the structures of DLA alleles were crystal structures and can be easily accessed by their unique PDB I.Ds.

Table 6. Epitope selection based on Epitope prediction score and antigenicity score.

Protein	DLA-Type	Epitope	Position of epitopes on the protein sequence	Epitope prediction Score	VaxiJen Antigenicity Score (Threshold value: 0.7)	Selection
VIRAL REPLICASE	DLA-8803401	FMFPHMYNY	295	0.8249350	0.5493	Reject
		ALYRRLTLI	273	0.5142250	0.7123	Select
		RVFITSNRL	251	0.4589590	0.0434	Reject
		KVWDGGNFI	282	0.3829500	-0.4171	Reject
CAPSID PROTEIN	DLA-8803401	YQHLPPFRF	86	0.4774550	0.9208	Select
		YIRAKWINW	101	0.3914190	2.2087	Select
VIRAL REPLICASE	DLA-8850101	FMFPHMYNY	295	0.6891070	0.5493	Reject
		RVFITSNRL	251	0.5503590	0.0434	Reject
		KVWDGGNFI	282	0.5206210	-0.4171	Reject

CAPSID PROTEIN	DLA-8850101	SVIQGANMV	212	0.4310240	0.5916	Reject
VIRAL REPLICASE	DLA-8850801	FMFPHMYNY	295	0.7916770	0.5493	Reject
		AAYPETYVK	140	0.7143770	0.3805	Reject
		GTMNFVARR	243	0.6422870	0.7786	Select
		HLQGFVNLK	60	0.5264080	0.8062	Select
		KVWDGGNFI	282	0.4437560	-0.4171	Reject
CAPSID PROTEIN	DLA-8850801	FIYDPLQDR	160	0.6825130	-0.0399	Reject
		HLPPFRFFK	88	0.6528390	-0.2234	Reject
		TLMENVLGR	112	0.6319340	-0.7453	Reject
		KLFHLRLRR	31	0.6040340	0.6741	Reject
		VQYDISAYI	249	0.5071050	0.1127	Reject

5.1.3. Molecular Docking and MD simulation

Docking studies were conducted to determine the efficiency of binding or interaction between DLA alleles and respective epitopes. Most negative ACE values and binding energies indicate perfect interaction between the selected epitopes and DLA alleles (Table 7). The 5 best possible docked complexes showed perfect hydrogen bond interactions during visualization analysis by deploying PyMol visualizing tool. All the H-bond interactions were shown in Figure 2. The bonding criteria for 5 selected complexes were found, as follows: 5F1I-GTMNFVARR (2.8Å Hydrogen bond Arg192 to Val6), 5F1I-HLQGFVNLK (3.2Å Hydrogen bond Ala79 to Val6), 7CJQ-ALYRRLTLI (3.1Å Hydrogen bond Ser57 to Thr7), 7CJQ- YIRAKWINW (2.2Å Hydrogen bond Glu129 to Trp9), 7CJQ- YQHLPPFRF (2.7Å Hydrogen bond Lys48 to Phe7). With the help of the Gromacs tool, molecular dynamics (MD) simulations were performed using the OPLS-AA field. OPLS-AA counts all atoms of docked complexes. The complex was stabilized towards the edges in a cubical box having a distance of 1.5 nm applying a clear TIP4P water model. Charges were neutralized by adding Na+ ions in this biomolecular simulation. The steepest descent technique of 200,000 steps with a 0.001 nm initial step size was employed for energy minimization. For MD simulations, a leap-frog algorithm to assimilate Newton's equations was used. Using LINCS (LINear Constraint Solver) algorithm and periodic boundary conditions PBC in all directions, the bond lengths were maintained. Cut-off of 0.9 nm distance and Particle-Mesh Ewald (PME) methods were used for the short-range interactions and long-range interactions respectively. The temperature (300 K) and pressure (1 bar) were regulated by a V-rescale thermostat and Parrinello-Rahman

barostat respectively. Molecular dynamics used in the isobaric-isothermal ensemble calculations without position restraint during 100 ns. Root mean square deviation (RMSD) plots indicate the selectable range of 0 to 8 Å for all considered epitopes in Figure 3.

Table 7. DLA-Epitope complex docking analysis: ACE values and Binding energy values for best possible complexes.

DLA-EPITOPE DOCKED COMPLEX	ATOMIC CONTACT ENERGY	BINDING ENERGY (Kcal/mol)
7CJQ- YQHLPPFRF	-147.66	-7.5
7CJQ- YIRAKWINW	-55.05	-6.8
7CJQ-ALYRRLTLI	-107.87	-7.8
5F1I- HLQGFVNLK	-168.42	-8.8
5F1I- GTMNFVARR	-64.97	-7.1

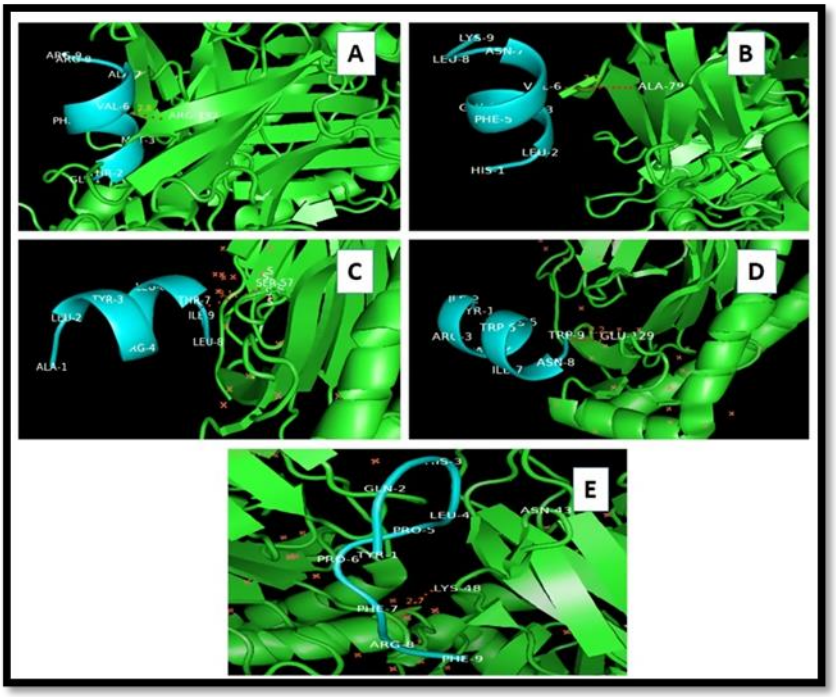


Fig 2. Docked complexes: **A)**5F1I-GTMNFVARR (2.8ÅHydrogen bond Arg192 to Val6) **B)**5F1I-HLQGFVNLK (3.2Å Hydrogen bond Ala79 to Val6) **C)**7CJQ-ALYRRLTLI (3.1Å Hydrogen bond Ser57 to Thr7) **D)** 7CJQ- YIRAKWINW(2.2Å Hydrogen bond Glu129 to Trp9) **E)** 7CJQ- YQHLPPFRF (2.7Å Hydrogen bond Lys48 to Phe7).

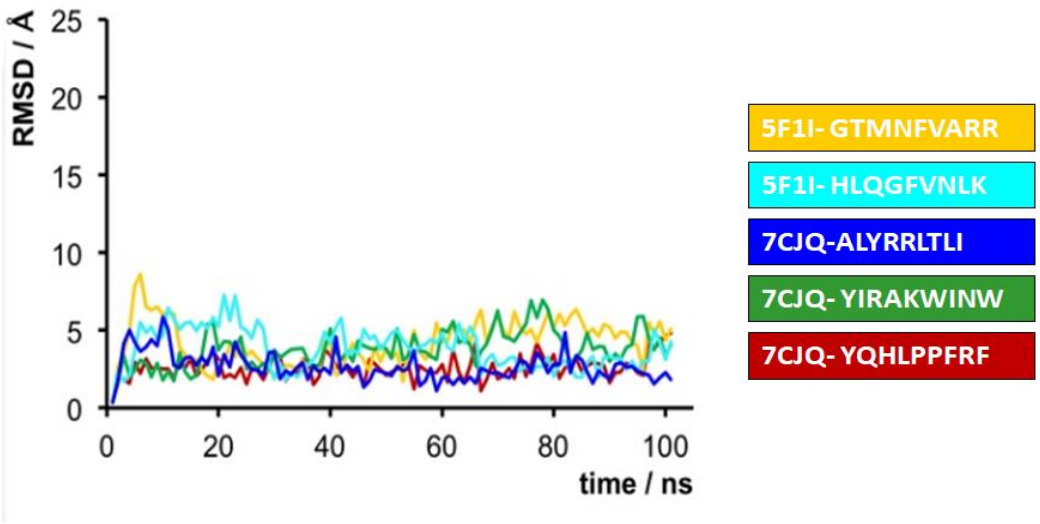


Fig. 3. RMSD plot for 5 selected epitopes-DLA complexes.

5.1.4 Full-fledged vaccine construction

The epitopes that were predicted to be conserved among different Canine circovirus strains, non-toxic, antigenic, and can induce the production of interferon-γ were chosen for the design of the final vaccine. RS09 and N and C terminal of *Salmonella Dublin* flagellin protein (UNIPROT ID: Q06971) were used as adjuvants. PADRE sequence was incorporated in the vaccine to develop stability and linked to each other by the GGS linker sequence. The final vaccine sequence is summarized in **Table 8.** Then, the physiochemical properties like stability, number of amino acids, molecular weight, isoelectric point, aliphatic index, number of positively and negatively charged amino acids, and various other properties and also the solubility of the vaccine were evaluated by the Solpro summarized in **Table 9**.

Table.8. The full-fledged vaccine sequence.

Final vaccine sequence	MAQVINTNSLSLLTQNNLNKSQSALGTAIERLSSGLRINSAKDDAAGQAIAN RFTANIKGLTQASRNANDGISIAQTTEGALNEINNNLQRVRELAVQSANSTN SQSDLDSIQAEITQRLNEIDRVSGQTQFNGVKVLAQDNTGGSAPPHALSGGS GTMNFVARRGGSAKFVAAWTLKAAAGGSHLQGFVNLKGGSAKFVAAWTL KAAAGGSALYRRLTLIGGSAKFVAAWTLKAAAGGSYIRAKWINWGGSYQH LPPFRFGGSLGNTVNNLTSARSRIEDSDYATEVSNMSRAQILQQAGTSVLAQ ANQVPQNVLSLLR

Table. 9. Physiochemical Properties of the full-fledged vaccine

Physicochemical characteristics	Values/ Description
Number of amino acids	320
Molecular weight	33544.51

Instability index	37.63 (Stable)
Aliphatic index	86.16
Theoretical pI	10.41
Extinction coefficient	33460
Total number of negatively charged residues (Aspartate + Glutamate)	17
Total number of positively charged residues (Arginine + Lysine)	30
Estimated half-life	30 hours in mammalian reticulocytes, >20 hours in yeast, and > 10 hours in *E. coli*
Total number of atoms	4712
Grand average of hydropathicity (GRAVY)	-0.225
Solubility determined using SOLpro server	Soluble (0.510037 probability)
Allergenicity	Non-allergen
Antigenicity using Vaxijen sever	Antigen (Vaxijen score:0.4892)

5.1.5. Predicted secondary and tertiary structure

The PSIPRED computed secondary structure of the vaccine contains beta sheets, alpha helices, and coils visualized in Figure 4. The I-TASSER predicted best tertiary structure of the vaccine model had a C-score of -0.60 and RMSD is 7.7±4.3A. C-score should be in the range of -5 to 2. The 3D structure of the vaccine construct is shown in **Figure 5**. The MOLprobity validated the quality of the tertiary structures of proteins and the Ramachandran plot is shown in **Figure 6.** 90.9% of all the residues were in the allowed or favored region.

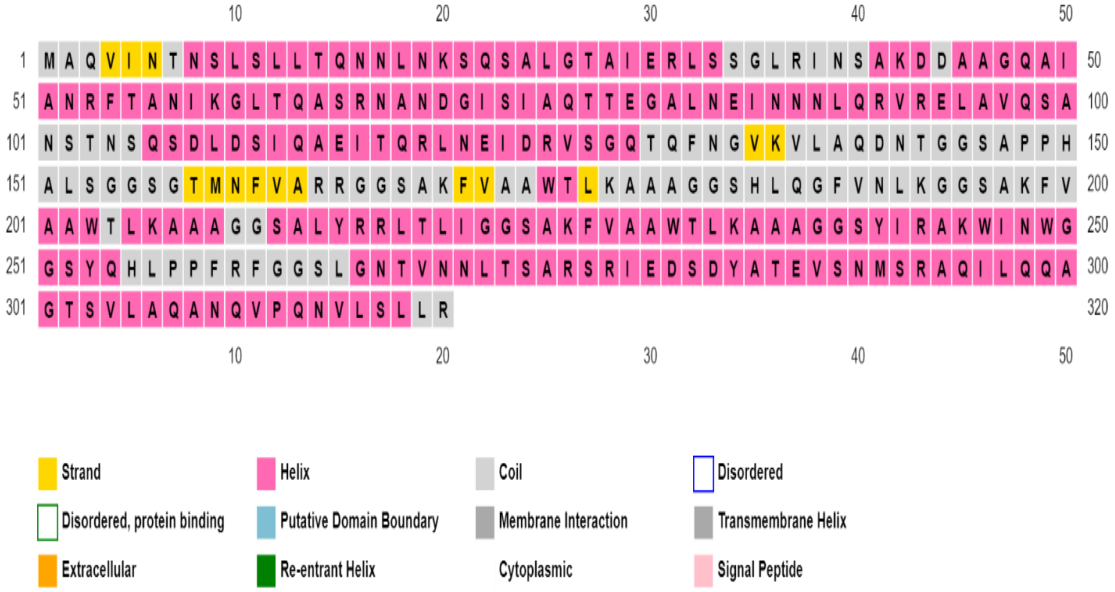


Fig. 4. Full-fledged vaccine sequence and secondary structure description.

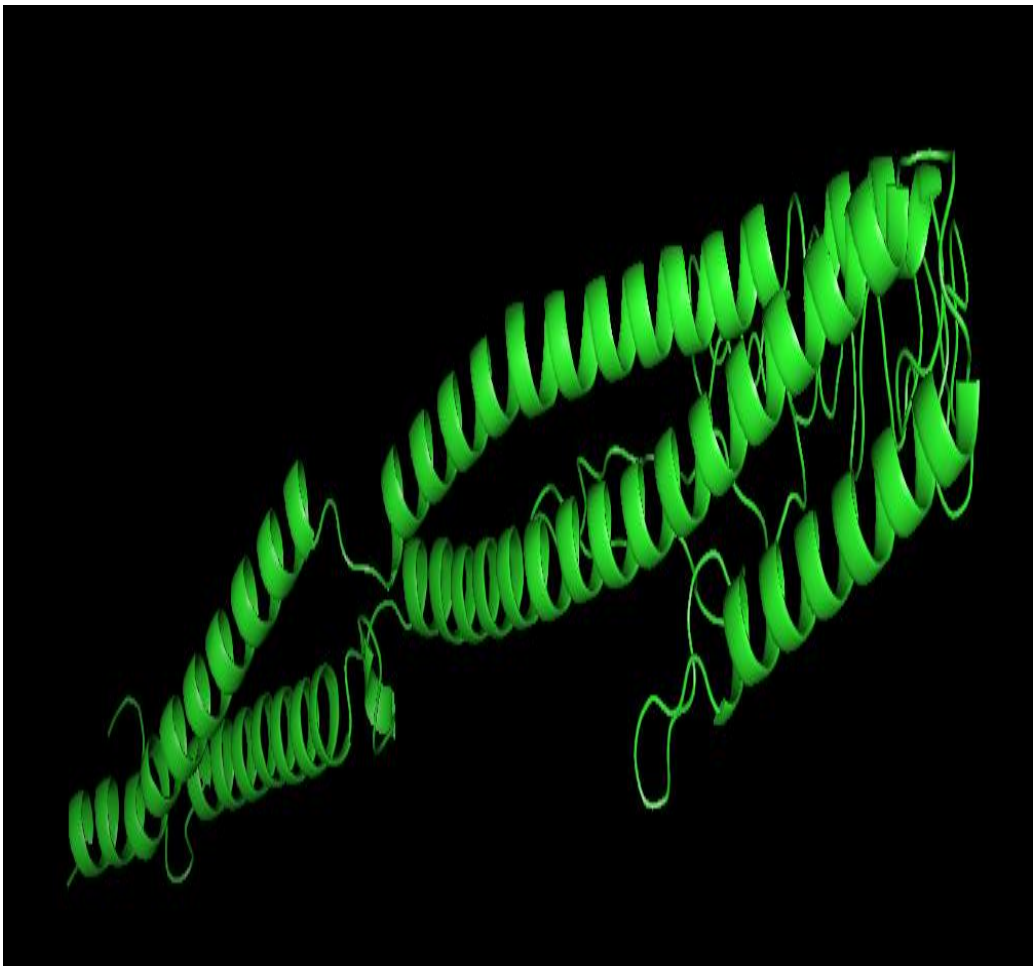


Fig.5. The 3D structure of vaccine construct.

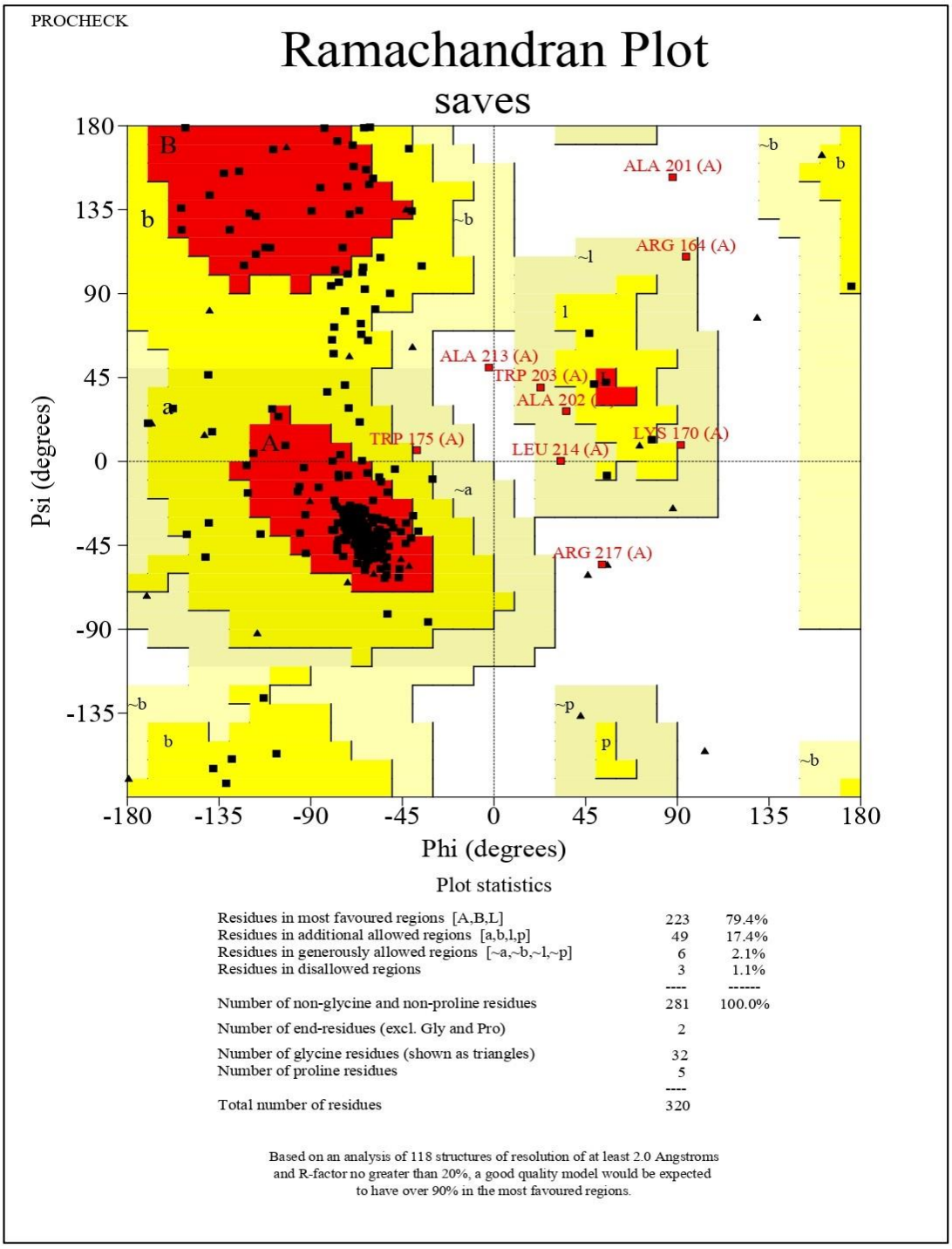


Fig. 6. Ramachandran plot for a full-fledged vaccine against Canine circovirus.

5.1.6 Codon Adaptation and vaccine *in-silico* cloning

For efficient cloning and expression of the vaccine in expression vectors, codon optimization is imperative. The codon-optimized JCAT vaccine construct had a CAI value of 0.98 and a GC content of 52.81. In **Table 10**, CAI and genetic inspectional set for a full-fledged vaccine is provided. During cloning, the codon-optimized

vaccine sequence was inserted between XhoI (158) and EcoRI(192) restriction sites of pET28a(+) vector as represented in **Figure 7**.

Table.10. Codon adaptation index for the gene of a finalized full-fledged vaccine against Canine circovirus.

CodonUsage adapted to Escherichia coli (strain K12)	
Improved DNA	
ATGGCTCAGGTTATCAACACCAACTCTCTGTCTCTGCTGACCCAGAACAA	**50**
CCTGAACAAATCTCAGTCTGCTCTGGGTACCGCTATCGAACGTCTGTCTT	**100**
CTGGTCTGCGTATCAACTCTGCTAAAGACGACGCTGCTGGTCAGGCTATC	**150**
GCTAACCGTTTCACCGCTAACATCAAAGGTCTGACCCAGGCTTCTCGTAA	**200**
CGCTAACGACGGTATCTCTATCGCTCAGACCACCGAAGGTGCTCTGAACG	**250**
AAATCAACAACAACCTGCAGCGTGTTCGTGAACTGGCTGTTCAGTCTGCT	**300**
AACTCTACCAACTCTCAGTCTGACCTGGACTCTATCCAGGCTGAAATCAC	**350**
CCAGCGTCTGAACGAAATCGACCGTGTTTCTGGTCAGACCCAGTTCAACG	**400**
GTGTTAAAGTTCTGGCTCAGGACAACACCGGTGGTTCTGCTCCGCCGCAC	**450**
GCTCTGTCTGGTGGTTCTGGTACCATGAACTTCGTTGCTCGTCGTGGTGG	**500**
TTCTGCTAAATTCGTTGCTGCTTGGACCCTGAAAGCTGCTGCTGGCGGTT	**550**
CTCATCTGCAGGGCTTCGTTAACCTGAAGGGTGGTTCTGCTAAATTCGTT	**600**
GCTGCTTGGACCCTGAAAGCTGCTGCTGGTGGTTCTGCTCTGTACCGTCG	**650**
TCTGACCCTGATCGGTGGTTCTGCTAAATTCGTTGCTGCTTGGACCCTGA	**700**
AAGCTGCTGCTGGTGGTTCTTACATCCGTGCTAAATGGATCAACTGGGGT	**750**
GGTTCTTACCAGCACCTGCCGCCGTTCCGTTTCGGTGGTTCTCTGGGTAA	**800**
CACCGTTAACAACCTGACCTCTGCTCGTTCTCGTATCGAAGACTCTGACT	**850**
ACGCGACTGAAGTAAGCAACATGTCTCGTGCTCAGATCCTGCAGCAGGCT	**900**
GGTACCTCTGTTCTGGCTCAGGCTAACCAGGTTCCGCAGAACGTTCTGTC	**950**
TCTGCTGCGT	
CAI-Value of the improved sequence:	**GC-Content of the improved sequence:**
0.9802705978765375	52.81249999999999

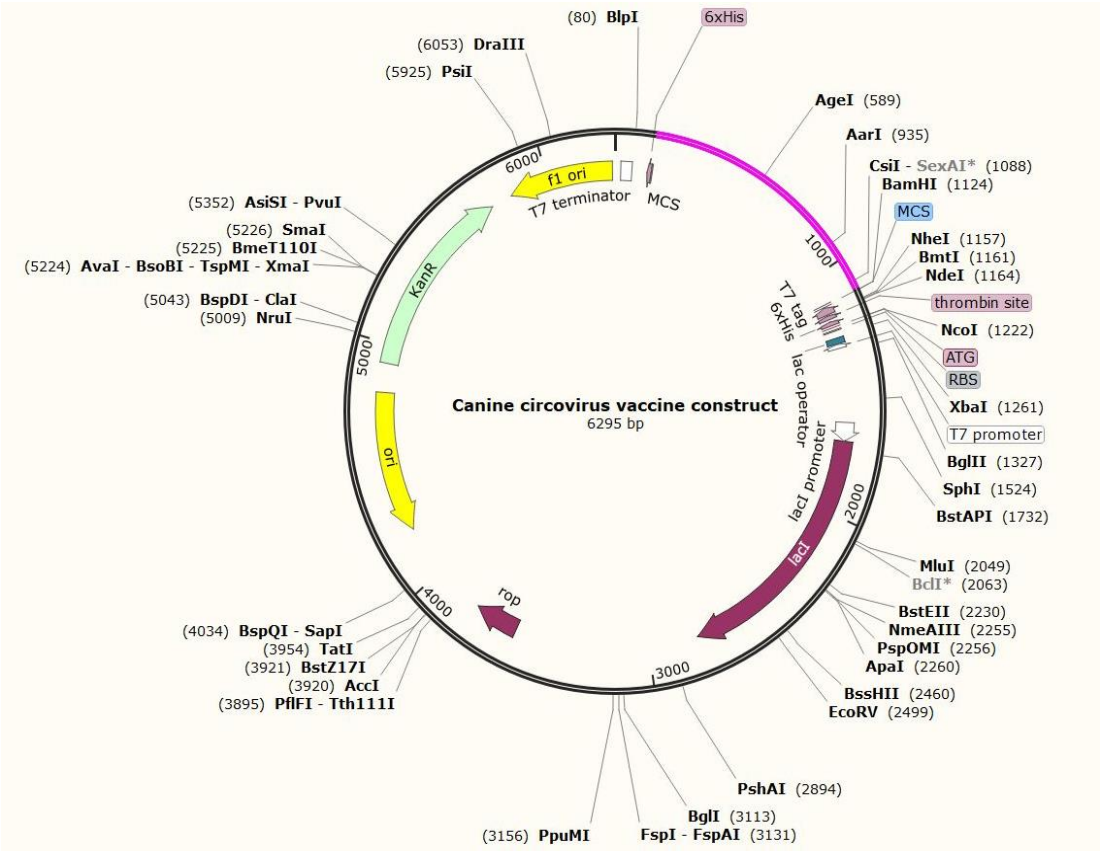


Fig. 7. Vaccine constructs in pET28a(+) vector: SnapGene restriction cloning module developed by Insightful Science was used for in silico cloning. During cloning the codon-optimized vaccine sequence was inserted between XhoI (158) and EcoRI (192) restriction sites of the pET28a(+) vector. (Vaccine construct is shown in magenta color)

There are two key inversely organized ORFs for *Circoviruses* that encode the replicase protein and capsid protein respectively **(Kapoor et al., 2012, Li et al., 2013)**. The replication-associated protein (Rep), which has sequence motifs characteristic of proteins involved in rolling circle replication **(Iliyana and Koonin., 1992)**, is the most conserved circovirus protein and is involved in the replication of the virus. In SARS-CoV, it has been reported that the replicase protein plays an important role in viral pathogenesis apart from its importance in the viral replication process **(Graham et al., 2008)**. As replicase protein is highly conserved and plays an important role in pathogenesis, it can be a good candidate to predict epitopes for designing a vaccine construct. The capsid-associated protein (Cap), is the structural component of the viral capsid and is the main antigenic protein of *Canine*

circoviruses because it uses repeated subunits to compose the entire capsid structure of the virus, Cap is significantly divergent and is characterized by an N-terminal region rich in basic amino acids that may provide DNA binding activity **(Niagro Niagro et al., 1998)**. Previously, three epitopes NKPWH, QSLFFF, and KHSRYFT, were predicted from the capsid protein of Porcine circovirus type3 PCV-3 **(Jiang et al., 2020)**. These epitopes were highly conserved B-cell epitopes and could be used to develop a vaccine against PCV-3 **(Jiang et al., 2020)** In another study the capsid protein has been targeted to develop a vaccine against the dengue-2 virus. This capsid-based vaccine developed cell-mediated immunity in monkeys. In this study, these two proteins of Canine circovirus: replicase and capsid proteins were targeted to design a multivalent epitope-based vaccine by immunoinformatics approach. The immunoinformatics approach has been previously used to design vaccine candidates against other animals, fish, and poultry pathogens such as *Avian avulavirus*, Foot and mouth disease virus, Seven banded grouper nervous necrosis virus, and *Mycoplasma galli septicum* **(Joshi et al., 2021)**. However, there has been no report of the development of a vaccine candidate against *Canine circovirus* by targeting its capsid and replicase protein by immunoinformatics approach so far. The *in silico* immunoinformatics approach has become the first step in developing vaccine candidates because of its cost-effective and time-saving manner. This approach can also help to overcome limitations such as genetic variations, antigenic shifts, and antigenic drifts during the vaccine development process **(Oyarzún and Kobe, 2016).** ViPR database was used for gathering information regarding viral proteome. Specific viral replicase and capsid proteins FASTA sequences retrieved from the NCBI database. The retrieved protein sequence data were analyzed by the NetMHCpan-4.1 webserver. The database also includes information about the antigens of dog leukocytes (Dog leukocyte antigen DLA) **(Reynisson et al., 2020).** Then the epitopes which could bind strongly with DLA molecules were evaluated for their antigenic potential by the Vaxijen webserver. The epitopes which had a Vaxijen score above 0.7 were selected for further analysis as they are considered to be highly antigenic. Altogether 5 epitopes were identified as highly antigenic: YQHLPPFRF, YIRAKWINW, ALYRRLTLI, HLQGFVNLK, and GTMNFVARR. Then, the structure of these antigenic epitopes was predicted. The interaction of these epitopes with the DLA alleles was evaluated by a molecular docking study. The negative binding energy between the interactions of epitopes with the DLA molecules implies

stable interactions and good binding affinity. Furthermore, the molecular dynamics simulations analysis also corroborated that the interactions among the epitopes and the DLA alleles were stable. After the selection of the epitopes, they were linked by GGS linkers to adjuvants such as RS09 (APPHALS) and N and C terminal sequence of *Salmonella dublin* flagellin protein along with PADRE sequence to design the final vaccine construct. These linkers and adjuvants have been previously used in multivalent epitope-based vaccine design of *Herpes Simplex Virus* and *Human Papilloma Virus, Candidaauris, Human Cytomegalovirus,* and *Dengue virus* **(Hasan et al., 2019; Krishnan G et al., 2021; Negahdaripour et al., 2017; Yazdani et al., 2020)**.PADRE (AKFVAAWTLKAAA) sequence was added to enhance the stability and potency of the vaccine construct **(Hasan et al., 2019)**. The physiochemical properties like stability, the number of amino acids, molecular weight, isoelectric point, aliphatic index, the number of positively and negatively charged amino acids, and various other properties of the vaccine were also determined by the ExPASy ProtParam web server **(Gasteiger et al., 2005).** The final vaccine construct has 320 amino acids and it was predicted to be stable, soluble, antigenic, and non-allergic by different web servers. To determine the secondary structure of the vaccine PSIPRED webserver was used **(Buchan and Jones, 2019)**. This web server predicted the presence of coils, strands, and helix in the final Canine circovirus vaccine construct. The tertiary structure of the vaccine was determined by using the I-TASSER server**.** The analysis of the tertiary structure by Ramachandran plot showed that the quality of the 3D model of vaccine construct was good. Most of the residues (90.9%) were in the allowed or favored region. The quality of the vaccine construct was also analyzed by C-score. Usually, the C- score should be between the range of -5 to 2 **(Yang et al., 2014)**. Higher C-score implies a good quality of the tertiary structure of the protein **(Yang et al., 2014)**. Here, the C-score of the vaccine construct was predicted as -0.60 which implies that the protein tertiary structure was of good quality. Finally, the *in-silico* cloning analysis showed that the vaccine can be cloned and expressed in Escherichia coli for commercial large production. The results obtained in this study so that the vaccine construct could be a good candidate in protecting dogs from Canine circovirus by generating a robust immune response. However, these claims need further validation by performing different *in vitro* and *in vivo* experiments.

5.2. In-silico Epitope based vaccine prediction and wet-lab validations

5.2.1 Prediction of Antigenic Epitopes and Analysis of their Antigenicity

NetMHCII Pan 4.0 server was deployed for screening the epitopes from capsid and replicase from four chosen strains of CanineCV. Our analyses resulted in the identification of 256 epitopes from capsid protein and 289 epitopes from replicase protein for each strain of CanineCV. Based on the VaxiJen score (with a threshold value of ≥0.7), 100 epitopes were screened out from 256 epitopes of capsid protein while 80 epitopes were screened from 289 epitopes of replicase protein for each strain. As a next step, we used the Genscript Optimum Antigen design tool and we found 9 antigenic epitopes from the capsid protein and 11 antigenic epitopes from the replicase protein of the CanineCV. Finally, we chose those epitopes that were commonly identified using two independent approaches: (a) NetMHCII Pan 4.0, **(Table 11)** and (b) Genscript Optimum antigen design tool **(Table 12)**. The resulting five epitopes that were conserved among all the four chosen strains were used to generate a multi-epitope vaccine construct. For the replicase proteins, the epitopes QVDQRGRDSRRGNP and LGGRGHFEPARGDD were conserved among all chosen strains. However, the epitopes RVRRHARASRRSYR, DPLQDRSSSRSFNM, and IAFKEFDYETGRQL were found conserved for the capsid in all the four CV strains considered in this study. Both capsid and replicase proteins in all the chosen strains were also subjected to multiple sequence alignment (MSA) which revealed identity to be greater than 94% between protein sequences. Since, Genscript Optimum Antigen and NetMHCII Pan 4.0 Server, both screen similar epitopes and based on the VaxiJen score; these epitopes can be considered antigenic and also conserved. For the final MEV construct design, only those epitopes selected were predicted as non-toxic, highly conserved, non-allergic, and antigenic by both Vaxijen and Genscript Optimum Antigen tools.

Table 11: Epitope screening based on VaxiJen score (> 0.7 Threshold) from NetMHCIIPan. 4.0 server

Protein Name	Epitopes	Protein ID	VaxiJen Score	Antigenicity
Replicase	QVDQRGRDSRRGNP	AVT56110.1 AWN93250.1 ALG63389.1 QFU80922.1	1.5960	ANTIGEN
	LGGRGHFEPARGDD	AVT56110.1 AWN93250.1 ALG63389.1 QFU80922.1	1.1867	ANTIGEN
Capsid	RVRRHARASRRSYR	AVT56111.1 AWN93251.1 ALG63390.1 QBQ20241.1	1.0516	ANTIGEN
	DPLQDRSSSRSFNM	AVT56111.1 AWN93251.1 ALG63390.1 QBQ20241.1	0.8144	ANTIGEN
	IAFKEFDYETGRQL	AVT56111.1 AWN93251.1 ALG63390.1 QBQ20241.1	1.0997	ANTIGEN

Table 12. List of the antigenic epitopes based on the Genscript optimum antigen tool and analysis of their toxicity, allergenicity, interferon-γ inducing ability, and conservancy with the selected epitopes are in bold and italics.

Protein ID	Peptide*	Vaxijen score	Antigen/ Non-antigen	Allergenicity	Toxicity	Interferon-γ inducing ability	Conservancy
QBQ20241.1 (capsid)	CDGEDQGRGNAQRSH	0.2792	Non-antigen	Non-allergen	Non-toxin	Negative	Non-conserved
	CVPGRLEPPKDPNK	-0.1227	Non-antigen	Non-allergen	Non-toxin	Negative	Non-conserved
	CQMKDMRPTTPDTST	0.7384	Antigen	Allergen	Non-toxin	Negative	Conserved in only three strains
	APVKPTNDPQTETPC	0.6359	Antigen	Non-allergen	Non-toxin	Negative	Conserved in only two strains
	CSGFKRGLTPKPMFT	0.1038	Non-antigen	Non-allergen	Non-toxin	Negative	Conserved in all four strains
	CIAFKEFDYETGRQL	1.251	Antigen	Non-allergen	Non-toxin	Negative	Conserved in all four strains

	CDPLQDRS SSRSFNM	0.7415	Antigen	Non-allergen	Non-toxin	Negative	Conserved in all four strains
	RVRRHARA SRRSYRC	1.3046	Antigen	Non-allergen	Non-toxin	Negative	Conserved in all four strains
	YQHLPPFR F	0.9208	Antigen	Non-allergen	Non-toxin	Negative	Non-conserved
QFU80922. 1 (replicase)	***QVDQRGR DSRRGNPC***	1.7055	Antigen	Non-allergen	Non-toxin	Negative	Conserved in all four strains
	CESGEVSR QGKRNDL	0.9999	Antigen	Non-allergen	Non-toxin	Negative	Conserved in only one strain
	LGGRGHF EPARGDDC	0.9514	Antigen	Non-allergen	Non-toxin	Negative	Conserved in all four strains
	CSPEMTTP RNWKTEV	-0.0674	Non-antigen	Allergen	Non-toxin	Negative	Non-conserved
	CDRYPLRV ETKGGT	0.8317	Antigen	Non-allergen	Non-toxin	Negative	Conserved in only three strains
	CNRLPHE WYSDEIGN	0.019	Non-antigen	Non-allergen	Non-toxin	Positive	Non-conserved
	PTPEEEEA VKNLAPC	0.4023	Antigen	Non-allergen	Non-toxin	Negative	Non-conserved
	CGCGKSRY CMETAPD	0.014	Non-antigen	Non-allergen	Non-toxin	Positive	Non-conserved
	SNKDYCSK GGDILIC	0.4868	Antigen	Non-allergen	Non-toxin	Negative	Conserved in only two strains
	ICGREVGE NGTPHLC	-0.0113	Non-antigen	Non-allergen	Non-toxin	Negative	Conserved in all four strains
	HLQGFVNL K	0.8062	Antigen	Allergen	Non-toxin	Negative	Conserved in all four strains

*An extra cysteine is added to N-or C- terminus for the identified epitopes to facilitate conjugation while predicting the antigens using the Optimum Antigen design tool.

5.2.2 Canine CV Vaccine Formulation and Determination of its Physiochemical Properties

The MEV construct design consisted of 372 amino acid residues. The vaccine included five immunogenic T cell epitopes; three adjuvants: RS09 having sequence APPHALS; PADRE having sequence AKFVAAWTLKAA, and an N-terminal and C-terminal sequence of *Salmonella typhimurium* flagellin protein that were linked to the epitopes with the help of GPGPG linker as an immune-adjuvant to elicit a robust immune response **(Duthie et al., 2011, Forstneric et al., 2017)**. It was shown that the flagellin's C terminal domain was required for TLR-5 activation **(Duthie et al., 2011,**

Forstneric et al., 2017). The sequence of formulated MEV construct is shown in Figure 8. The selected five epitopes were CIAFKEFDYETGRQL, CDPLQDRSSSRSFNM, RVRRHARASRRSYRC, QVDQRGRDSRRGNPC, and LGGRGHFEPARGDDC which were predicted to be conserved in all the selected strains and were highly antigenic, non-toxic and non-allergic, **Table 12**. The formulated vaccine was antigenic with an antigenic score of 0.5089 as calculated by the VaxiJen 2.0 server. The final vaccine had a molecular weight of 38.44 kDa and an iso-electric point (pI) of 9.91. The total numbers of negative and positive charged amino acids were 27 and 39, respectively. The calculated aliphatic index (69.14) indicated a stable vaccine construct in a broad range of temperatures. Our vaccine had an estimated half-life of 30h in mammalian reticulocytes (*in vitro*) and with>20 hours (in *Sacchormyces cervisiae*, *in vivo*) and >10 hours (*Escherichia coli*, *in vivo*) as analyzed by the Expasy ProtParam server. Finally, the predicted solubility of the vaccine was calculated to be 0.661 upon overexpression in *E. coli*.

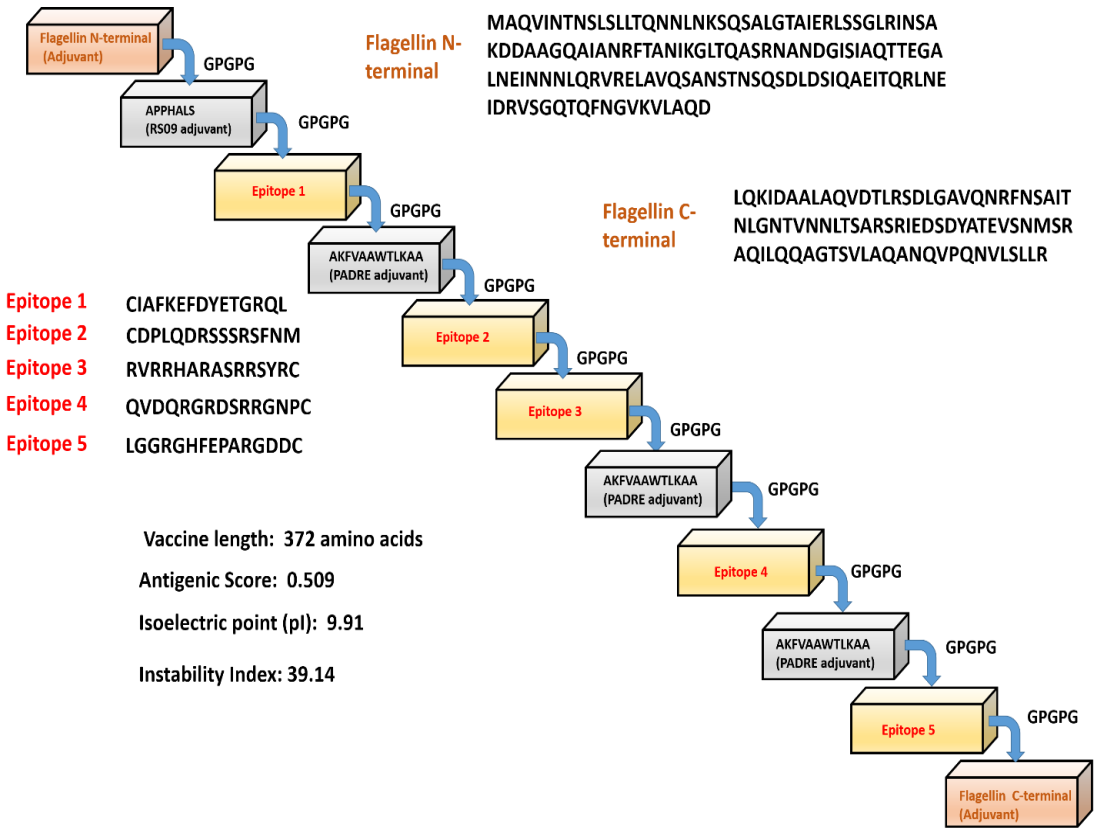


Fig. 8: Schematic representation of the predicted multi-epitope vaccine (MEV) construct for CanineCV with different adjuvants, linkers, and epitopes with their respective amino acid sequence.

5.2.3 Modelling and Docking of TLR-5 and Multi-epitope Vaccine Construct

The secondary structure of the vaccine construct shows the presence of mostly helix and coils in the MEV construct. The tertiary structure of the final multi-epitope vaccine construct and the immunogenic TLR-5 were predicted using I-TASSER program **(Yang et al., 2014)**. To confirm the quality of predicted structures, we calculated the Ramachandran plot of the modeled TLR-5 and vaccine construct. For the multi-epitope vaccine construct, out of 372 residues, 76.6% of residues were found to fall in the core-favoured regions, and 20.3% and 1.4% of residues were in allowed and generously allowed regions, and the remaining 1.7% residues were observed under disallowed regions, see Figure 9a. However, for TLR-5 receptor, 73% of amino acids were found in the core acceptable region and the other 24.5% were observed under the allowed region and with an additional 1.5% falling under the generously allowed region and the remaining 1% in the disallowed regions of Ramachandran plot see Figure 9b.

The molecular docking of a modeled structure of a multi-epitope vaccine construct with the modeled TLR-5 from *Canis lupus* was performed using the HADDOCK 2.4 webserver to evaluate the interaction between the vaccine construct and TLR-5 which would probably be eliciting an immune response **(Chawla et al., 2021)**. The data-driven docking of the designed multiepitope-TLR5 complex was performed. The flagellin adjuvant which is a bacterial protein is known to specifically bind to TLR5 and illicit the innate immune response **(Forstneric et al., 2017)**. Here, we specifically used information-driven docking where the information about which specific residues are involved in the interaction is needed to drive the docking calculations. **Jacchieri et al., 2003** used a similar approach while using complementary hydropathy between the sequences of flagellin and TLR-5 to predict the potential binding sites and the structure of complex. The sequences 'LQRVRELAVQ' and 'EILDISRNQL' were predicted as the potential binding sites in flagellin and human TLR5, respectively. However, for *Canis lupis* family "EVLDVSRNQL" sequence was used due to slight amino acid variations at some positions for canine TLR-5 as a hot spot for potential binding studies. The vaccine construct was docked with TLR-5 receptor using HADDOCK program which resulted in clustering 129 structures in 11 cluster(s), which represents 64% of the water-refined HADDOCK generated models. We had specifically chosen the top-ranked cluster with the lowest HADDOCK score which is the most significant one for docking analysis. The HADDOCK score of -96.4 ± 5.1

indicated a good interaction between the vaccine construct and TLR-5 receptor. The score of buried surface area (BSA) was 1977.3 ± 54.4 Å^2 which reflects proximity and a less water-exposed region on the protein binding surface. F**igure 10** illustrates the molecular docking between vaccine construct and TLR-5 receptor. Our docking simulations predicted the structure where Epitope 2 and Epitope 3 were present in proximity to the TLR-5 receptor. In particular, the ionic interactions were present between Arg210 from Epitope 2 and Asp232 from TLR-5, and Arg222 from Epitope 3 and Asp318 from TLR-5r, see Figure 10. Further, the hydrophobic interaction was present between Phe212 from Epitope 2 and Val259 from TLR-5. To study the stability of the docked complex, the docked complex was subjected to MD simulations, see below for detailed analysis.

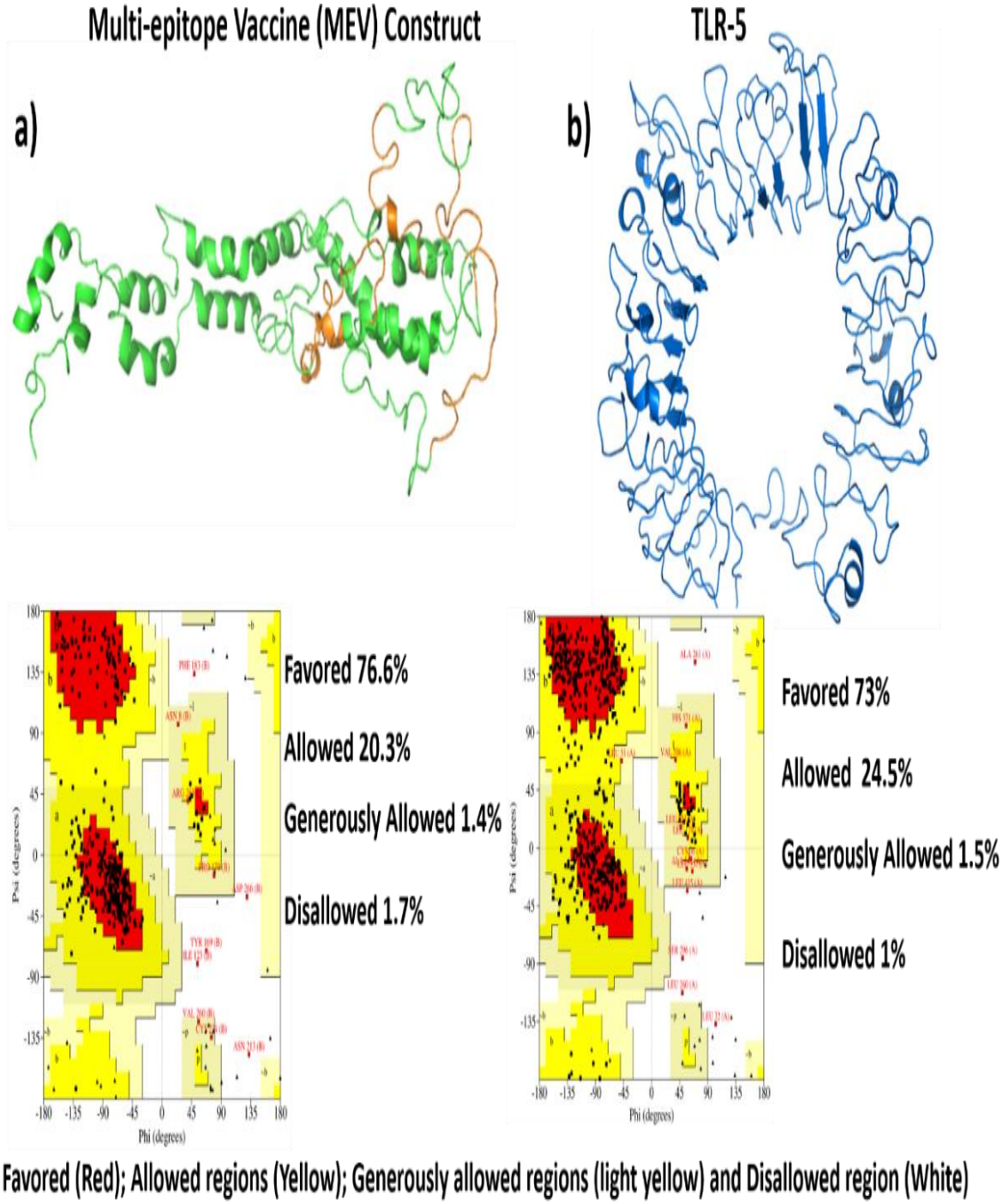


Fig. 9: The modeled 3D structures of the a) MEV construct. The five epitopes are shown in orange and the bacterial flagellin sequence is present at the N- and C-terminal of the vaccine construct and is colored in the green cartoon. b) Cartoon representation of a modeled Toll-like Receptor 5 (TLR-5) used in this study. Respective Ramachandran plots are also shown with predicted structures.

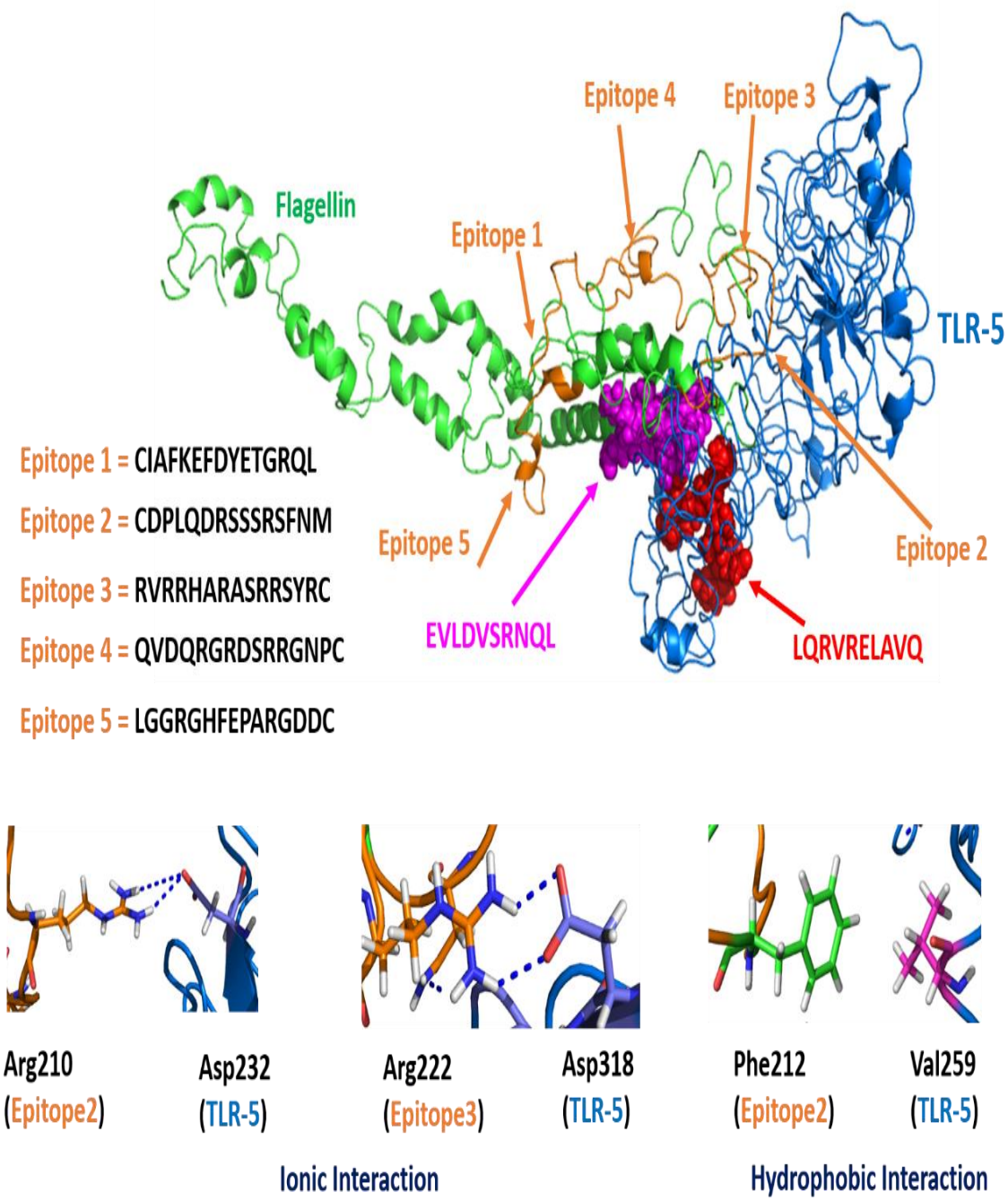


Fig. 10: Upper Panel) The docked complex of MEV construct (green cartoon) and TLR-5 (marine blue cartoon). The hotspot residues used (see Methods section for details) for information-driven docking are shown in spheres and colored in magenta and red for MEV construct and TLR-5 receptor respectively. The epitope peptide sequence present in MEV is also shown; Lower Panel). Three representative interactions between immunogenic receptor (TLR-5) amino acid residues and amino acids from epitopes 2 and 3 from MEV construct.

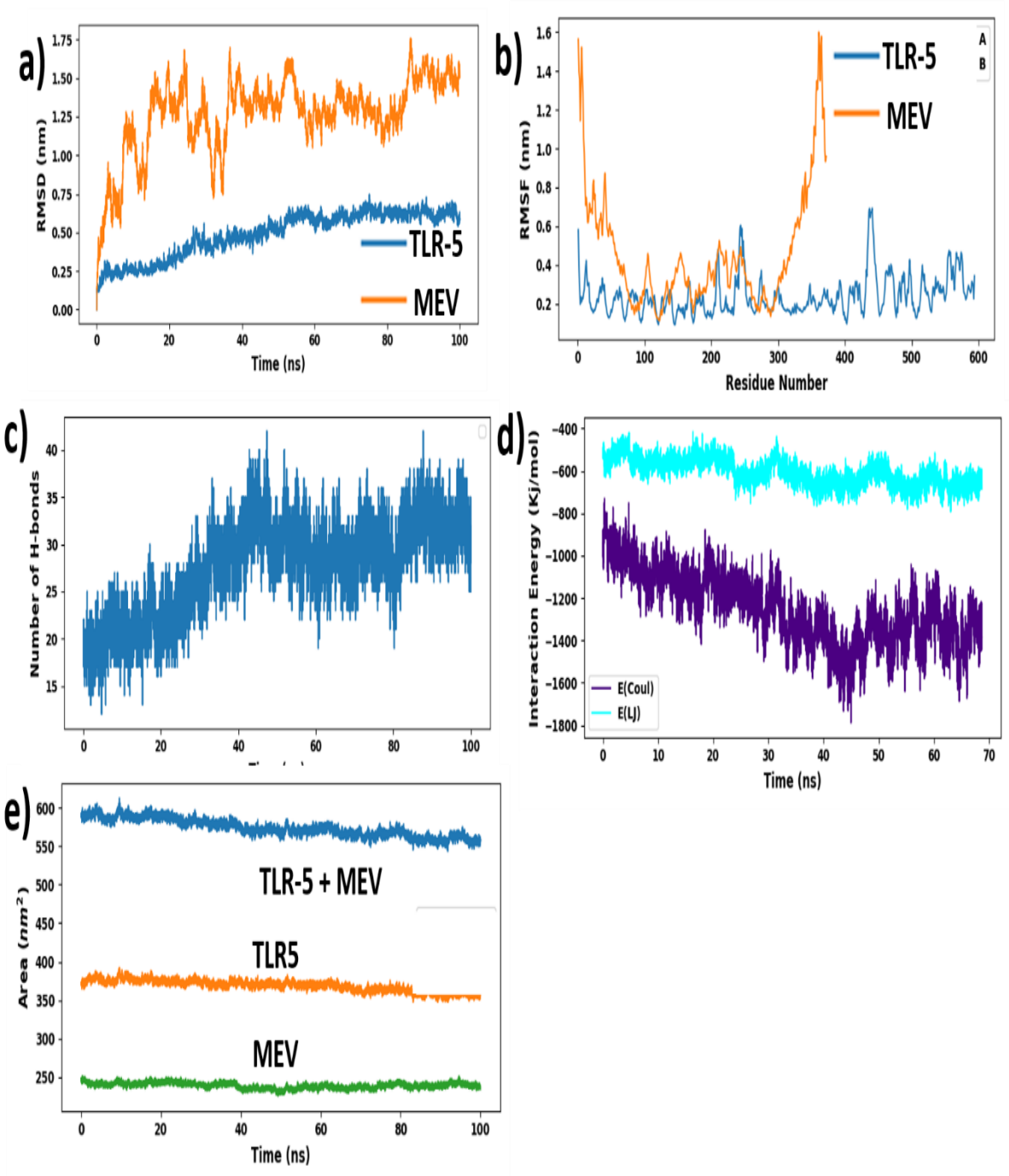


Fig. 11: a) Time evolution of backbone RMSD of TLR-5 and MEV construct during MD simulations; b) Backbone RMSF plots; c) Time evolution of the number of hydrogen bonds between TLR-5 and MEV construct; d) Variation in Interaction energy plot with segregated electrostatics (E-Coul) and van der Waals (E-LJ) component. e) Solvent accessible surface area of TLR-5 and MEV construct.

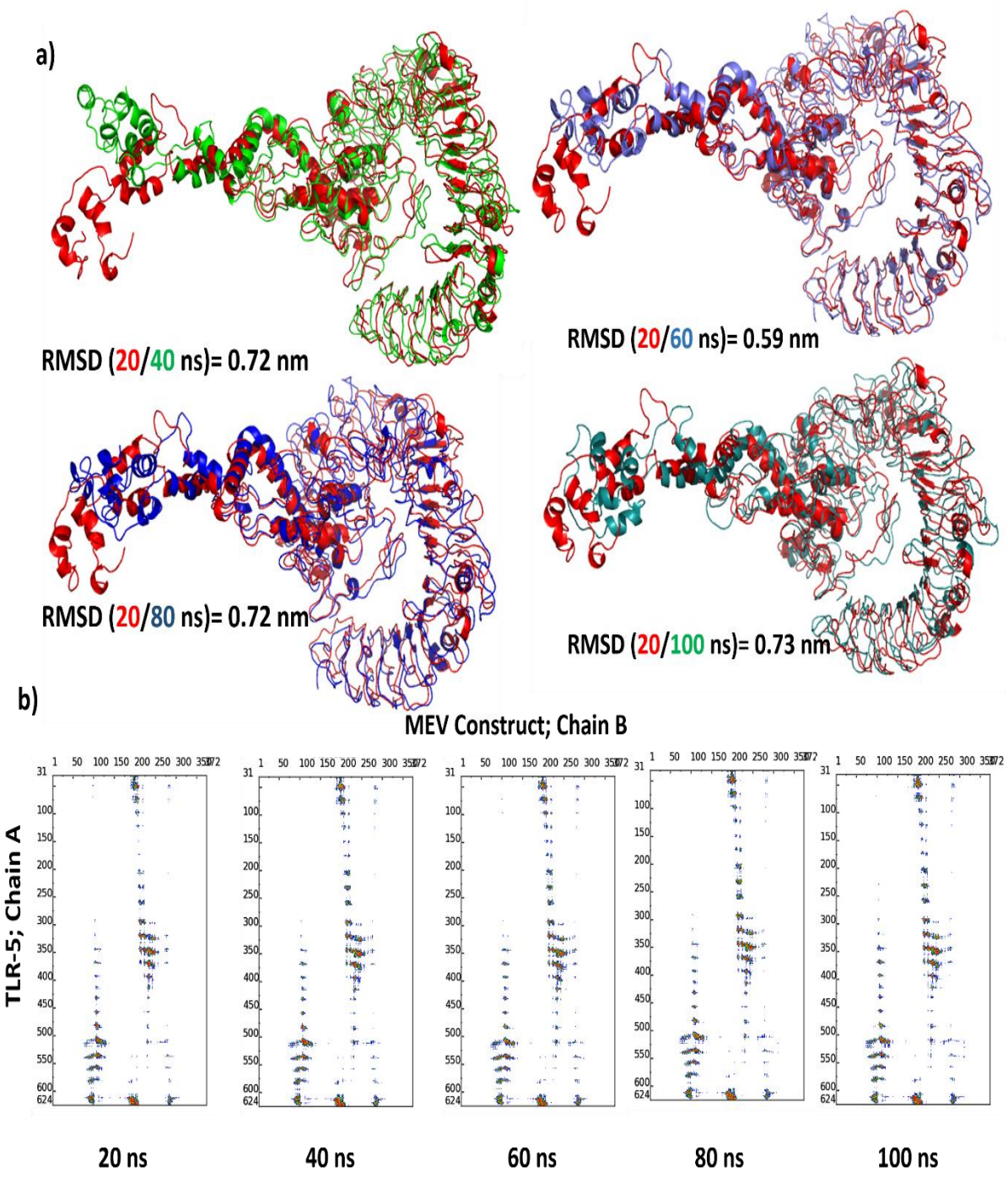


Fig12: a) Superimposition of selected snapshots of TLR-5 and MEV construct and their respective RMSD values; b) Contact maps showing the conservation of contacts between residues in between TLR-5 and Vaccine construct. The dots at the crossover of two residues are colored in red, yellow, green, and blue if any pair of an atom is closer than 7, 10, 13, and 16 Å.

5.2.4 Structural Stability of TLR-5 and MEV complex

To study the structural stability of the docked complex of MEV and TLR-5 MD simulations were performed **(Karplus and Kuriyan., 2005)**. Here, the GROMACS software was used for MD simulation studies to understand the structural properties and interactions between the docked MEV and TLR-5 complex. In total, a simulation of 100ns was performed for TLR-5 docked with MEV construct. Further, different structural parameters were analyzed with the first being the Root Mean Square Deviation (RMSD) which is indicative of stability and conformational drift of MEV and TLR-5 complex in the simulation. In particular, the RMSD of the Cα atoms from their initial position of TLR-5 and MEV was monitored as a function of simulation time and was plotted in Figure11a. The calculated RMSD value of the docked TLR-5 and MEV complex after 40 ns and after 100ns simulations were 0.72 and 0.73 nm, respectively. The highest RMSD between the initial conformation and conformation obtained after 100ns simulation time was evident from the flexible terminal region of the flagellin moiety of the MEV construct. The TLR-5 receptor alone was substantially stable see Figure 11a, with an average RMSD value of 0.49±0.15nm. However, the predicted multi-epitope vaccine possessed flexibility due to the presence of epitopes and linkers regions that formed the loops indicating a relatively high RMSD value of 1.26±0.25 nm. However, looking at Figure 11a, it was clear that the RMSD of the multi-epitope vaccine got stabilized after 40ns of simulation time. Next, the Root Mean Square Fluctuation (RMSF) parameter was plotted separately for both the TLR-5 and vaccine constrict, where the highly flexible regions were found associated with elevated fluctuations. Looking at Figure 11b, it was evident that the calculated RMSF values for Cα atoms of TLR-5 remained stable for the overall structure. In contrast, a high fluctuation was observed for some regions with the residues ranging from 1-90 (the region associated with N-terminal flagellin in MEV construct) and amino acids 312-372 (regions associated with C-terminal of flagellin in MEV construct); indicating the flexible regions in MEV construct with an average RMSF of 1.13± 0.6 nm. Further, the number of hydrogen bonds between TLR5 and MEV construct complex were analyzed which remained constant after 20ns of the simulations, see Figure 11c. It is illustrated in Figure 11e that the large solvent accessible surface area was observed for the TLR-5 complexed with MEV, as compared to the isolated TLR-5 and MEV construct. Finally, we calculated the

interaction energies between TLR5 and MEV construct and further decomposed the interaction energy component into coulombic interactions (E Coul), which represents the interaction accounting for the electrostatics between TLR-5 and MEV construct. The other energy component is Lennard Jones (E(LJ)) potential, which accounts for the van der Waals interactions between TLR-5 and MEV. Considering the complex remained stable after 30ns of simulation time, the interaction energy components were plotted for the final 70 ns. From figure 11 d, it was evident that the predominant interaction that stabilized the overall TLR-5 and MEV complex stemmed from the electrostatic component of the interaction, with the average electrostatic energy of -1250.57±167.7Kcal/mol. These interactions could play a major role in the binding of MEV with TLR-5. In contrast, the average E(LJ) contribution was found to be -608.68±63.5kcal/mol. Thus, our overall energetic analysis depicted that the predominant stabilization was contributed by the electrostatic component of the interaction energy between the interface region of TLR-5 and the MEV construct.

5.2.5 Detailed Analysis of Snapshots obtained during MD Simulations and its Comparison with Initial Modeled TLR5-MEVConstruct.

We carried out a detailed analysis of the non-covalent interactions occurring between the MEV and TLR-5 using an in-house script. The docked TLR-5 and MEV construct complex was stabilized by a network of 60H-bonding interactions within 3.5Å and 13 hydrophobic interactions and 11 salt-bridge interactions within 5Å and 6Å distance cut-off, respectively. In fact, out of a total of 11 salt bridge interactions, two existed between Epitope 2 of vaccine construct and TLR-5, and three existed between Epitope 3 of vaccine constrict and TLR-5. The distance plots of the nearest C-alpha atoms for the salt-bridge interactions indicated the stability of these interactions throughout the simulation time. This further reinforced the stability of the predicted vaccine construct and TLR-5 immune receptor. To further assess the stability of the overall complex, the superimposition of structures at different time steps was done which resulted in a good overlap between structures. From figure 12a, it was clear that the complex remained stable with a low RMSD for the selected snapshots. Moving forward, the interface analysis for selected snapshots was performed between TLR-5 and vaccine docked construct using the COCOMAPS tool **(Vangone et al., 2012, Vangone et al., 2011)**, see figure 12b. Contact maps were computed, see figure 12b, where the dots at the crossover of two residues were colored in red, yellow, green, and

blue if any pair of atoms between vaccine construct and TLR-5 molecule were closer than 7, 10, 13, and 16 Å. Looking at figure 12b, it was evident that the overall contacts remained stable for the selected snapshots. The interface area is 2001 $Å^2$, with a percentage of polar and non-polar residues at the interface of approximately 58% and 42% respectively. Using a cut-off distance of 5Å to define two atoms in contact between vaccine construct and TLR-5, the predominant contacts existed between hydrophilic-hydrophilic (79 in number) and hydrophilic-hydrophobic (71 in number) residues and which was followed by 39 contacts between hydrophobic-hydrophobic residues.

To access the stability of TLR-5 complexed with the MEV construct, the conservation of inter-residue contacts (ICs) at the interface along the simulation time was plotted using the MD cons program **(Abdel-Azeim et al., 2014)**. The ICs at the interface indicated the conservation of different contacts as a measure of similarity between different snapshots. The consensus map of 1000 MD snapshots was plotted in figure 13a, with the selected stable ICs distances plotted in figure 13b. The overall conservation of the residue sat the interface during the MD simulations could be visually appreciated by looking at the MD consensus map, see figure 13a. From the MD consensus plot, it was clear that many contacts between the MEV construct and TLR-5 remained stable throughout the simulation time. MD cons analysis resulted in C_{70} value of 0.64, which indicates that 64% of ICs were conserved in at least 70% of frames. Thus, the conservation of many contacts at the interface during the simulation is suggested. A few examples of intermolecular contacts that remained stable over the simulation time between TLR-5 and epitope 2 and epitope 3 of the MEV complex are illustrated in figure 13b. In particular, the salt-bridge interaction existed between Asp393 of TLR-5 and Arg230 from Epitope3 of the MEV construct, which remained stable along the simulation time with an average distance between Asp393(CB)-Arg230(CZ) of 0.40±0.02nm. Next, a stable hydrogen bonding interaction between Thr619 of TLR-5 and Asp201 from Epitope2 of MEV construct was observed that remained stable with an average distance of Thr619(CB) and Asp201(CG) to be 0.47±0.06 nm. Further, hydrophobic contact between Ile50 from TLR-5 and Pro202 from Epitope2 of MEV construct with an average distance of Ile50(CB) and Pro202(CB) atom to be 0.39±0.06 nm was also observed to be stable along the simulation time. MD cons analysis of selected snapshots indicated that the overall

interaction pattern at the interface between TLR-5 and MEV construct remained stable throughout the simulation time.

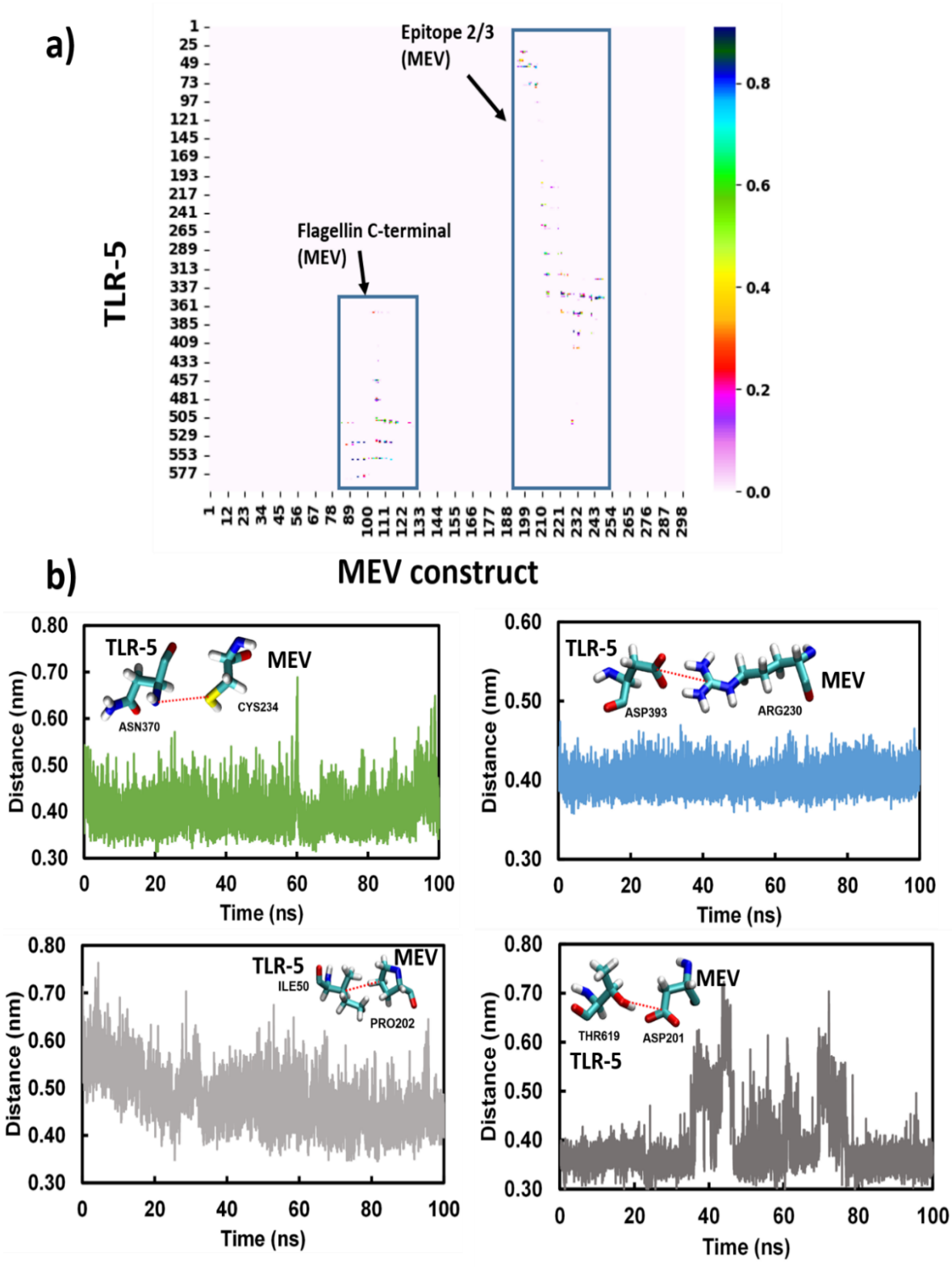


Figure 13: MD cons consensus map of 1000 MD snapshots. a) The region of Epitope 2/3 and C-terminal of flagellin of MEV that interacts with TLR-5 is highlighted in the blue box. The ICs that were most conserved are shown in distance plots; b) Four examples of conserved contacts (indicated by dark blue dots) between TLR-5 and Vaccine construct.

5.2.6 Peptide Synthesis and Conjugation to Carrier Protein

GenScript's microwave-based PepPower™ technology was used to synthesize the selected capsid protein epitope RVRRHARASRRSYRC. The cysteine residue was added to the N-terminus of the immunogenic peptide. The peptide antigen containing cysteine is indicated as CRVRRHARASRRSYRC which was subsequently conjugated with KLH.

5.2.7 *In-vivo* Validation for Antibody Development and Assessment

The affinity-purified concentration of the antibody is found to be 287µg/ml of the antiserum of New Zealand white rabbits subcutaneously immunized with KLH-conjugated peptide. To confirm whether the KLH conjugate peptide can stimulate humoral immunity and produce antibodies, an indirect ELISA was performed to determine the antibody titer. The antibody titer is expressed as the inverse of the highest dilution of the antibody which gives a positive result in ELISA. The serum collected from rabbit is serially diluted more than 8 folds. Interestingly, 2-6-fold diluted aliquots (i.e., 1:1,000 to 1:64,000) show a positive result on ELISA, as indicated by the titer value of ≥2.28 with reference signal/blank ≥2.1; (see supplementary information file for more details). However, no colour is observed when IgG in the range of 1:1,000 to 1:512,000 are incubated with the antigen, see Figure 14. These results indicate that the vaccine designed in this study is a promising immunogen that could effectively stimulate the production of antibodies *in vivo*.

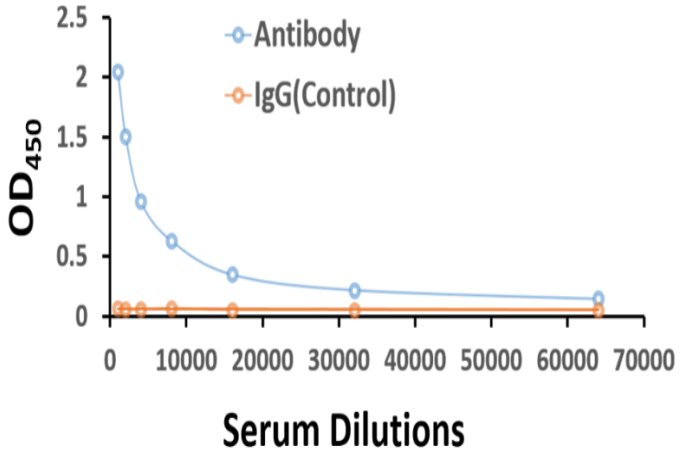


Figure 14: ELISA assay scatter plot of OD_{450} values and serum dilutions from (1:1000 to 1:64,000) of CanineCV antibody and IgG control with varying concentrations of antigen. The OD_{450} values can be inversely correlated to serum dilutions.

A multi-epitope peptide vaccine targeting multiple antigens approach has been extensively explored for vaccine development for animal, fish, and poultry viral diseases including avian avulavirus, foot and mouth disease virus, and seven banded grouper nervous necrosis virus(ref). However, no vaccine candidate is reported for CanineCV until now. Multi-epitope based vaccine-based prediction for CanineCV could assist in managing severe problems of gastroenteritis and diarrhoea in the Dogs' population. Further, it could be effectively successful in terms of saving time without conducting multiple random hit-and-trial approaches. Deploying this strategy, we have recently screened the proteins of CanineCV to predict the epitopes using the latest neural network algorithm-dependent tools like NetMHC server; along with antigenicity and toxicity analyses criteria to brief the major group of considerable epitopes during subsequent steps of screening **(Jain et al., 2021)**. In the present study, we extended this robust immunoinformatics approach to identify a novel $CD4^+$ T cell epitope-based MEV candidate for CanineCV taking into account four different strains of *Canis lupis* family from different countries. Further, we identified epitopes that were conserved in all the strains considered in the present work, and one of the potential epitopes was synthesised for further validation using *in vivo* technique. As a first step, we predicted 2180 MHCII binding $CD4^+$T cell epitope peptides. Further, we refined five conserved epitopes among all four strains of the CanineCV according to their antigenicity score. The antigenic and nontoxic epitopes

‘CIAFKEFDYETGRQL’,‘CDPLQDRSSSRSFNM’,‘RVRRHARASRRSYRC’,‘QVD QRGRDSRRGNPC’ and ‘LGGRGHFEPARGDDC’ that were conserved among all the four CanineCV strains were selected for final vaccine construct. Once the epitopes were identified, they were linked by GPGPG linkers to adjuvants such as RS09 and flagellin protein along with the PADRE sequence to craft the final vaccine construct. Subsequently, TLR-5 was selected as the receptor for MEV since TLR has been reported for its role in the internalization of vaccine construct which would probably elicit the immune response. Moreover, TLR-5 has been recently used for the successful internalization of the vaccine constructs for viral diseases including West Nile virus, HBV, influenza virus, and porcine circovirus. Overall, our computational analyses suggest that the novel MEV construct identified is not only a potent immunogen but also non-toxic and non-allergenic in nature. Further, the *in vivo* analysis of MEV ensured $CD4^{+}$T cell dependant generation of antibodies; thus, clearly indicating the vaccine’s potential to induce the immune response.

Antigen details for Canine Circovirus

Virus	**Protein**	**Identifier**	**Epitope**
Canine circovirus	Capsid	QBQ20241.1 capsid Protein	**RVRRHARASRRSYRC**

No	**Antigenic Determinant**	**Length**	**Antigenicity/ Surface/ Hydrophilicity**	**Disordered Score**	**Synthesis**	**Oryctolagus cuniculus blast**
1	RVRRHARA SRRSYRC	15	2.11/1.00/1.04	0.3643	Easy	46%

ELISA titer >1:128,000 with immunogen and ≥2mg purified antibody*2 (Purity>80%) can be guaranteed.

Procedure:

1. For toxicity analysis: in the mammalian cell line.

2. For Antibody development: This experiment was performed in rabbits. ELISA was also performed to determine the titer of the antibodies produced in rabbits against your peptides. Also, determine the antigenicity of the peptides using algorithms.

3. Animal immunization

4. Antibody purification

5. Custom Polyclonal Antibody Development Against Peptide Antigen

Sequence:

RVRRHARASRRSYRC

The blue regions are predicted as antigenic determinants.

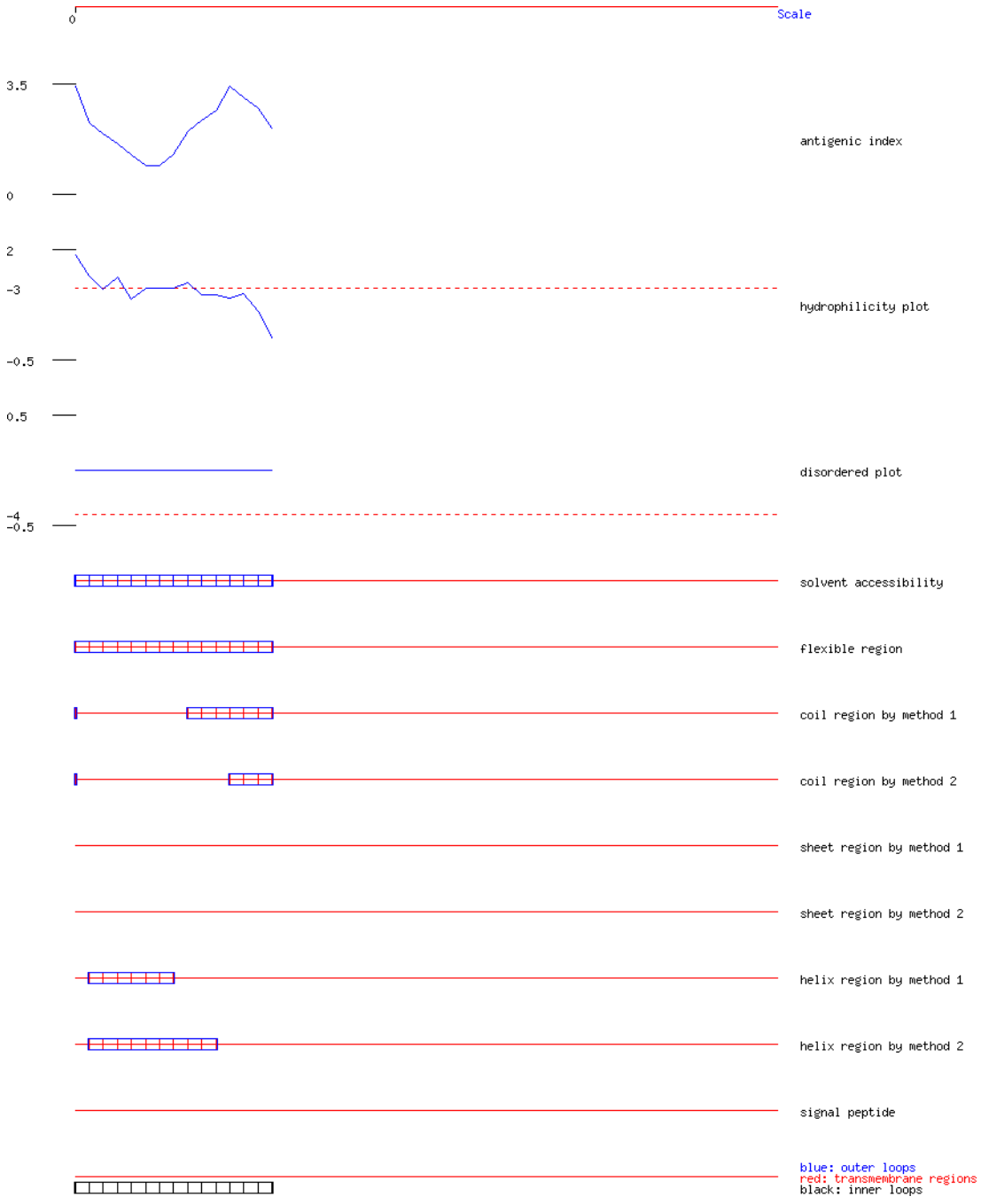
The underline regions are predicted as disordered regions.

The regions which background gray are predicted as transmembrane regions.

Protein Properties Image:

This Information graphics are based on OptimumAntigen™ design tool.

The cutoff of Antigenic Index: -3.00



Antigen Designed:

No	Start	Antigenic Determinant	Length	Antigenicity/Surface/ Hydrophilicity	Disordered Score	Synthesis	Oryctolagus_ cuniculus blast
1	1	RVRRH ARASR RSYRC	15	2.11/1.00/1.04	0.3643	Easy	46%

1. An extra "C" (high-lighted as green) is added to the C-terminus (or N-terminus) to facilitate conjugation.

2. Positively charged residues (K,R,H) are in blue. Negative charged residues (D,E) are in red.

3. For a basic peptide, initially try to dissolve the peptide in water; if the peptide does not dissolve, try 10% and higher solutions of acetic acid; if the peptide still does not dissolve, add TFA(<50ul) to solubilize the peptide and dilute to 1ml with deionized water.

For an acidic peptide, initially try to dissolve the peptide in water, if the peptide does not dissolve, add NH_4OH(<50ul) to solubilize the peptide and dilute to 1ml with deionized water.

4. Disorder peptide note: The linear conformation of a disordered peptide is more similar to its natural conformation in the folded protein.

ELISA protocol

Reagents

Coating Buffer (1XPBS Buffer)

8.5 g NaCl
1.4 g Na_2HPO_4
0.2 g NaH_2PO_4
1000 ml d H_20
Adjust pH to pH 7.4
Store at 4°C

Washing Buffer

0.5 ml Tween 20
1000 ml PBS Buffer
Store at 4°C

Blocking Buffer

100 ml Washing Buffer
1 g BSA
Store at 4°C

Stop Buffer

8.3 ml 12 mol/L HCl
91.7 ml ddH_2O
Store at 4°C

TMB Reagent (GenScript Cat.No M00078)

Indirect ELISA protocol

Procedure

Coating

19. Dilute the antigen with *Coating Buffer* and coat appropriate wells of ELISA plate with the antigen by pipeting 100 µl of the diluted solution. The concentration of coated antigen ranges from 1 -10 µg/ml.

20. Cover the plate with an adhesive plastic and incubate for 2 hours at 37^{o}C or 4^{o}C overnight.

21. Remove the antigen coating solution from the wells of plate by flicking the plate over a sink.

22. Wash the plate three times by adding the wells with 200 µl of *Washing Buffer*.

Blocking

23. Add 200 µl of *Blocking Buffer* to block the non-specific binding sites in the coated wells.

24. Cover the plate with an adhesive plastic and incubate for at 2 hours at 37^{o}C or overnight.

25. Remove the *Blocking Buffer* from the wells of plate by flicking the plate over a sink.

26. Wash the plate three times by adding the wells with 200 µl of *Washing Buffer.*

Incubation

Dilute the primary antibody or antiserum with *Blocking Buffer* and add 100 µl of the diluted antibody to each well of the plate. The concentration of primary antibody is based on the manufacturer's instructions

27. Cover the plate with an adhesive plastic and incubate for 1 hour at 37^{o}C or overnight at 4°C.

28. Remove the diluted primary antibody solution from the wells of plate by flicking the plate over a sink.

29. Wash the plate three times by adding the wells with 200 µl of washing *buffer*.

30. Dilute the HRP-conjugated secondary antibody with *Blocking Buffer* and add 100 µl of the diluted secondary antibody to each well of the plate.

31. Cover the plate with an adhesive plastic and incubate for 30 minutes at 37^{o}C.

32. Remove the diluted secondary antibody from the wells of plate by flicking the plate over a sink.

33. Wash the plate five times by adding the wells with 200 µl of *Washing Buffer.*

Detection

34. Add 100 µl of the *TMB Reagent* per well with a multichannel pipete or a multipipete.

35. After sufficient color development, add 100 µl of *Stop Buffer* to the wells. Note: 15-30 minutes is enough for color development.

36. Read the absorbance of each well with a plate reader.

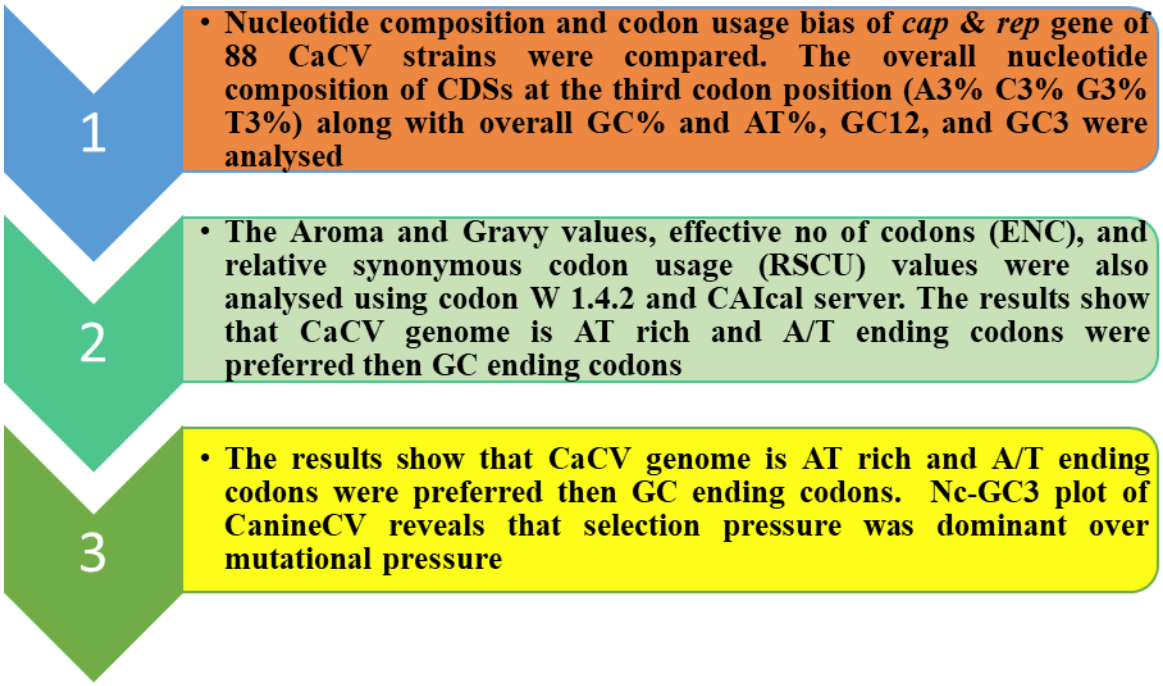
CHAPTER 6
SUMMARY & CONCLUSION

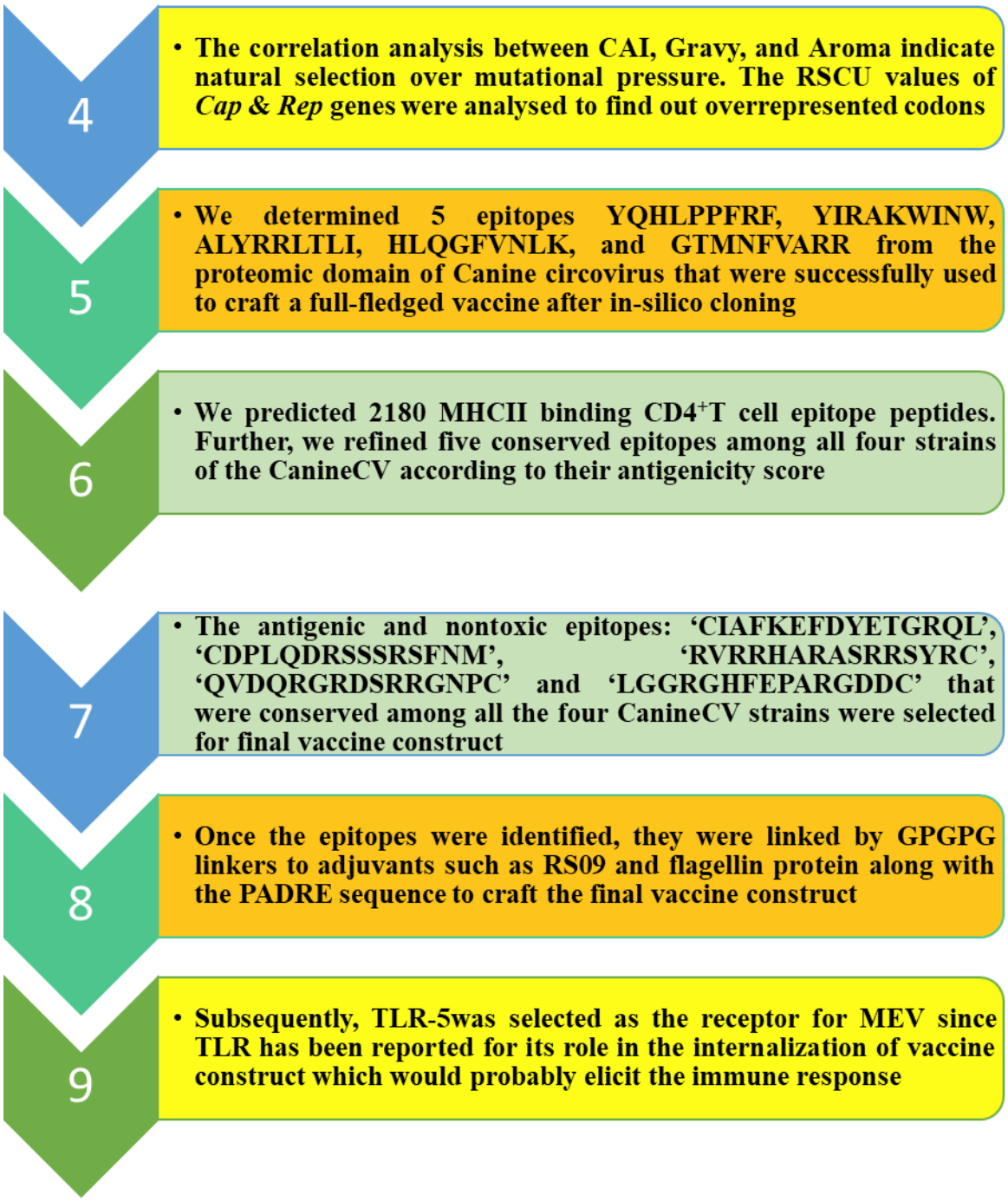
Nucleotide composition and codon usage bias of *cap* & *rep* gene of 88 CaCV strains were compared. The overall nucleotide composition of CDSs at the third codon position (A3% C3% G3% T3%) along with overall GC% and AT%, GC12, and GC3 were analysed. The Aroma and Gravy values, effective no of codons (ENC), and relative synonymous codon usage (RSCU) values were also analysed using codon W 1.4.2 and CAIcal server. The results show that CaCV genome is AT rich and A/T ending codons were preferred then GC ending codons. Nc-GC3 plot of CanineCV reveals that selection pressure was dominant over mutational pressure. The correlation analysis between CAI, Gravy, and Aroma indicate natural selection over mutational pressure. The RSCU values of *Cap* & *Rep* genes were analysed to find out overrepresented codons.

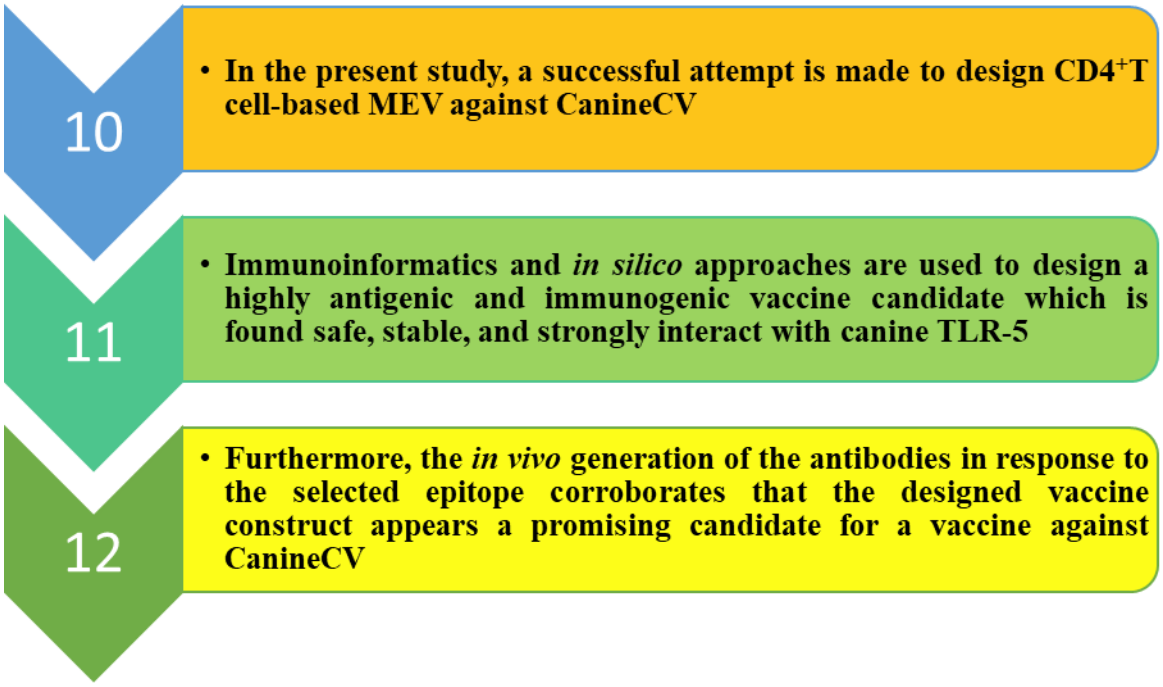
In this research, we designed epitope-based vaccine against Canine circovirus by deploying an immunoinformatics approach. We determined 5 epitopes YQHLPPFRF, YIRAKWINW, ALYRRLTLI, HLQGFVNLK, and GTMNFVARR from the proteomic domain of Canine circovirus that were successfully used to craft a full-fledged vaccine after in-silico cloning. This help in reducing time taking trial and error peptide screening for identifying vaccine candidates. This is a very cost-effective method and easy to conduct before wet-lab trials. Such methods open new dimensions in veterinary research. In the present study, we extended this robust immunoinformatics approach to identify a novel $CD4^+$ T cell epitope-based MEV (Multi-epitope vaccine) candidate for CanineCV taking into account four different strains of *Canis lupis* family from different countries. Further, we identified epitopes that were conserved in all the strains considered in the present work, and one of the potential epitopes was synthesised for further validation using *in vivo* technique. As a first step, we predicted 2180 MHCII binding $CD4^+$T cell epitope peptides. Further, we refined five conserved epitopes among all four strains of the CanineCV according to their antigenicity score. The antigenic and nontoxic epitopes: 'CIAFKEFDYETGRQL', 'CDPLQDRSSSRSFNM', 'RVRRHARASRRSYRC', 'QVDQRGRDSRRGNPC' and 'LGGRGHFEPARGDDC' that were conserved among all the four CanineCV strains were selected for final vaccine construct. Once the epitopes were identified, they were linked by GPGPG linkers to adjuvants such as RS09 and flagellin protein along with the PADRE sequence to craft the final vaccine construct. Subsequently, TLR-5 was selected as the receptor for MEV since TLR has

been reported for its role in the internalization of vaccine construct which would probably elicit the immune response. No MEV-based vaccine has been reported for CanineCV to date. In the present study, a successful attempt is made to design $CD4^{+}$T cell-based MEV against CanineCV. Immunoinformatics and *in silico* approaches are used to design a highly antigenic and immunogenic vaccine candidate which is found safe, stable, and strongly interact with canine TLR-5. Furthermore, the *in vivo* generation of the antibodies in response to the selected epitope corroborates that the designed vaccine construct appears a promising candidate for a vaccine against CanineCV.

SUMMARY FLOWCHART




4
• The correlation analysis between CAI, Gravy, and Aroma indicate natural selection over mutational pressure. The RSCU values of *Cap* & *Rep* genes were analysed to find out overrepresented codons
5
• We determined 5 epitopes YQHLPPFRF, YIRAKWINW, ALYRRLTLI, HLQGFVNLK, and GTMNFVARR from the proteomic domain of Canine circovirus that were successfully used to craft a full-fledged vaccine after in-silico cloning
6
• We predicted 2180 MHCII binding CD4+T cell epitope peptides. Further, we refined five conserved epitopes among all four strains of the CanineCV according to their antigenicity score
7
• The antigenic and nontoxic epitopes: 'CIAFKEFDYETGRQL', 'CDPLQDRSSSRSFNM', 'RVRRHARASRRSYRC', 'QVDQRGRDSRRGNPC' and 'LGGRGHFEPARGDDC' that were conserved among all the four CanineCV strains were selected for final vaccine construct
8
• Once the epitopes were identified, they were linked by GPGPG linkers to adjuvants such as RS09 and flagellin protein along with the PADRE sequence to craft the final vaccine construct
9
• Subsequently, TLR-5was selected as the receptor for MEV since TLR has been reported for its role in the internalization of vaccine construct which would probably elicit the immune response


10
• In the present study, a successful attempt is made to design CD4+T cell-based MEV against CanineCV
11
• Immunoinformatics and *in silico* approaches are used to design a highly antigenic and immunogenic vaccine candidate which is found safe, stable, and strongly interact with canine TLR-5
12
• Furthermore, the *in vivo* generation of the antibodies in response to the selected epitope corroborates that the designed vaccine construct appears a promising candidate for a vaccine against CanineCV

BIBLIOGRAPHY

1. Abdel-Azeim, S., Chermak, E., Vangone, A., Oliva, R. and Cavallo, L. (2014) MDcons: Intermolecular contact maps as a tool to analyze the interface of protein complexes from molecular dynamics trajectories. *Bmc Bioinformatics*, **15**. DOI: 10.1186/1471-2105-15-S5-S1.

2. Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J.C., Hess, B. and Lindahl, E. (2015) GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, **1-2**, 19-25.DOI: 10.1016/j.softx.2015.06.001.

3. Anderson, A., Hartmann, K., Leutenegger, C. M., Proksch, A. L., Mueller, R. S., & Unterer, S. (2017). Role of Canine circovirus in dogs with acute haemorrhagic diarrhoea. *Veterinary Record*, 180(22), 542-542. DOI:10.1136/vr.103926

3. Aravind, S., Kamble, N.M., Gaikwad, S.S., Khulape, S.A., Dey, S., Dhama, K. and Mohan, C.M., 2014. Bioinformatics study involving characterization of synonymous codon usage bias in the duck enteritis virus glycoprotein D (gD) gene. *Asian Journal of Animal and Veterinary* Advances,*9*(4), pp.229-242.
DOI: 10.3923/ajava.2014.229.242.

4. Athey, J., Alexaki, A., Osipova, E., Rostovtsev, A., Santana-Quintero, L.V., Katneni, U., Simonyan, V. and Kimchi-Sarfaty, C., 2017. A new and updated resource for codon usage tables. *BMC bioinformatics*, *18*(1), p.391. DOI:10.1186/s12859-017-1793-7.

5. Baha, S., Behloul, N., Liu, Z., Wei, W., Shi, R. and Meng, J., 2019. Comprehensive analysis of genetic and evolutionary features of the hepatitis E virus. *BMC genomics*, *20*(1), p.790. DOI:10.1186/s12864-019-6100-8.

6. Baker, S.F., Nogales, A. and Martínez-Sobrido, L., 2015. Downregulating viral gene expression: codon usage bias manipulation for the generation of novel influenza A virus vaccines. *Future virology*, *10*(6), pp.715-730. DOI: 10.2217/fvl.15.31.

7. Belalov, I.S. and Lukashev, A.N., 2013. Causes and implications of codon usage bias in RNA viruses. *PLoS One*, *8*(2). DOI: 10.1371/journal.pone.0056642.

8. Biswas, K.K., Palchoudhury, S., Chakraborty, P., Bhattacharyya, U.K.,

Ghosh, D.K., Debnath, P., Ramadugu, C., Keremane, M.L., Khetarpal, R.K. and Lee, R.F., 2019. Codon usage bias analysis of Citrus tristeza virus: Higher codon adaptation to citrus reticulatahost. *Viruses*, *11*(4), p.331. DOI:10.3390/v11040331.

9. Buchan, D. W. A., & Jones, D. T. (2019). The PSIPRED Protein Analysis Workbench: 20 years on. *Nucleic Acids Research*, 47(W1), W402–W407. DOI:10.1093/nar/gkz297.

9. Bulmer, M., 1987. Coevolution of codon usage and transfer RNA abundance. *Nature*, *325*(6106), pp.728-730. DOI:10.1038/325728a0.

10. Butt, A.M., Nasrullah, I., Qamar, R. and Tong, Y., 2016. Evolution of codon usage in Zika virus genomes is host and vector specific. *Emerging microbes & infections*, *5*(1), pp.1- 14. DOI:10.1038/emi.2016.106.

11. Castells, M., Victoria, M., Colina, R., Musto, H. and Cristina, J., 2017. Genome-wide analysis of codon usage bias in Bovine Coronavirus. *Virology journal*, *14*(1), p.115. DOI: 10.1186/s12985-017-0780-y.

12. Chawla, M., Gorle, S., Shaikh, A.R., Oliva, R. and Cavallo, L. (2021) Replacing thymine with a strongly pairing fifth Base: A combined quantum mechanics and molecular dynamics study. *Computational and Structural Biotechnology Journal* , **19**, 1312-1324.DOI: 10.1016/j.csbj.2021.02.006.

13. Chen, S.L., Lee, W., Hottes, A.K., Shapiro, L. and McAdams, H.H., 2004. Codon usage between genomes is constrained by genome-wide mutational processes. *Proceedingsof the National Academy of Sciences*, *101*(10), pp.3480-3485. DOI:10.1073/pnas.0307827100.

14. Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., & Richardson, D. C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D: Biological Crystallography*, *66*(1), 12-21. DOI: 10.1107/S0907444909042073.

15. Chen, Y., Sun, J., Tong, X., Xu, J., Deng, H., Jiang, Z., Jiang, C., Duan, J., Li, J., Zhou, P. and Wang, C., 2014. First analysis of synonymous codon usage in porcine circovirus. *Archives of Virology*, *159*(8), pp.2145-2151. DOI:10.1007/s00705-014-2015-5.

16. Coleman, J.R., Papamichail, D., Skiena, S., Futcher, B., Wimmer, E. and Mueller, S., 2008. Virus attenuation by genome-scale changes in codon pair bias. *Science, 320*(5884), pp.1784- 1787.DOI: 10.1126/science.1155761.

17. Cruz, T. F., Batista, T. N., Vieira, E. M., Portela, L. M. F., Baccarin, A. M., Gradiz, J. J., & Araújo Junior, J. P. (2020). Genomic characterization of Canine circovirus detected in a dog with intermittent hemorrhagic gastroenteritis in Brazil. *Ciência Rural*, 50(5), e20190909.DOI:10.1590/0103-8478cr20190909.

18. Darden, T., York, D. and Pedersen, L. (1993) Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. *Journal of Chemical Physics*, **98**, 10089-10092.DOI: 10.1063/1.464397.

19. Dash, R., Das, R., Junaid, M., Akash, M.F.C., Islam, A. and Hosen, S.Z., 2017. In silico- based vaccine design against Ebola virus glycoprotein. *Advances and applications in bioinformatics and chemistry: AABC, 10*, p.11.DOI: 10.2147/AABC.S115859.

20. David, A., Islam, S., Tankhilevich, E. and Sternberg, M.J.E. (2022) The AlphaFold Database of Protein Structures: A Biologist's Guide. *Journal of Molecular Biology*, **434**, 167336.DOI: 10.1016/j.jmb.2021.167336.

21. Decaro, N., Martella, V., Desario, C., Lanave, G., Circella, E., Cavalli, A., Elia, G., Camero, M. and Buonavoglia, C., 2014. Genomic characterization of a circovirus associated with fatal hemorrhagic enteritis in dog, Italy. *PloS one*, *9*(8). DOI:10.1371/jour- nal.pone.0105909.

22. Dhanda, S.K., Vir, P. and Raghava, G.P. (2013) Designing of interferon-gamma inducing MHC class-II binders. *Biology Direct*, **8**, 30. doi: 10.1186/1745-6150-8-30.

23. Doytchinova, I.A. and Flower, D.R. (2007) VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *Bmc Bioinformatics*, **8**. doi: 10.1186/1471-2105-8-4.

24. Du, J., Dungan, S.Z., Sabouhanian, A. and Chang, B.S., 2014. Selection on synonymous codons in mammalian rhodopsins: a possible role in optimizing

translational processes. *BMC evolutionary biology*, *14*(1), p.96. DOI: 10.1186/1471-2148-14-96.

25. Duthie, M.S., Windish, H.P., Fox, C.B. and Reed, S.G. (2011) Use of defined TLR ligands as adjuvants within human vaccines. *Immunological Reviews*, **239**, 178-196. DOI: 10.1111/j.1600-065X.2010.00978.x.

26. Fan, R.L., Valkenburg, S.A., Wong, C.K., Li, O.T., Nicholls, J.M., Rabadan, R., Peiris, J.M. and Poon, L.L., 2015. Generation of live attenuated influenza virus by using codon usage bias. *Journal of virology*, *89*(21), pp.10762-10773. DOI:10.1128/JVI.01443-15.

27. Forstneric, V., Ivicak-Kocjan, K., Plaper, T., Jerala, R. and Bencina, M. (2017) The role of the C-terminal D0 domain of flagellin in activation of Toll like receptor 5. *Plos Pathogens*, **13**. DOI: 10.1371/journal.ppat.1006574.

28. Franzo, G., Segales, J., Tucciarone, C.M., Cecchinato, M. and Drigo, M., 2018. The analysis of genome composition and codon bias reveals distinctive patterns between avian and mammalian circoviruses which suggest a potential recombinant origin for Porcine circovirus 3. *PloS one*, *13*(6). DOI:10.1371/journal. pone.0199950.

29. Fu, M., 2010. Codon usage bias in herpesvirus. *Archives of virology*, *155*(3), pp.391- 396. DOI:10.1007/s00705-010-0597-0.

30. Galtier, N., Roux, C., Rousselle, M., Romiguier, J., Figuet, E., Glémin, S., Bierne, N. and Duret, L., 2018. Codon usage bias in animals: disentangling the effects of natural selection, effective population size, and GC-biased gene conversion. *Molecular biology and evolution*, *35*(5), pp.1092-1103. DOI:10.1093/molbev/msy015.

31. Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. *The Proteomics Protocols Handbook*, 571–607. doi:10.1385/1-59259-890-0:571

32. Giraldo-Ramirez, S., Rendon-Marin, S., Vargas-Bermudez, D. S., Jaime, J., &

Ruiz-Saenz, J. (2020). First detection and full genomic analysis of Canine circovirus in CPV-2 infected dogs in Colombia, South America. *Scientific Reports*, *10*(1), 1-9. doi: 10.1038/s41598-020-74630-8.

33. Grantham, R., Gautier, C., Gouy, M., Mercier, R. and Pave, A., 1980. Codon catalog usage and the genome hypothesis. *Nucleic acids research*, *8*(1), pp.197-197. DOI:10.1093/nar/8.1.197c.

34. Grewal, R.K., Shaikh, A.R., Gorle, S., Kaur, M., Videira, P.A., Cavallo, L. and Chawla, M. (2021) Structural Insights in Mammalian Sialyltransferases and Fucosyltransferases: We Have Come a Long Way, but It Is Still a Long Way Down. *Molecules*, **26**. doi: 10.3390/molecules26175203.

35. Grote, A., Hiller, K., Scheer, M., Munch, R., Nortemann, B., Hempel, D. C., & Jahn, D. (2005). JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Research*, 33(Web Server), W526–W531. doi:10.1093/nar/gki376

36. Gu, W., Zhou, T., Ma, J., Sun, X. and Lu, Z., 2004. Analysis of synonymous codon usage in SARS Coronavirus and other viruses in the Nidovirales. *Virus research*, *101*(2), pp.155-161. DOI:10.1016/j.virusres.2004.01.006.

37. Gun, L., Yumiao, R., Haixian, P. and Liang, Z., 2018. Comprehensive Analysis and Comparison on the Codon Usage Pattern of Whole Mycobacterium tuberculosis Coding Genome from Different Area. *BioMed research international*. DOI:10.1155/2018/3574976.

38. Gupta, S., Kapoor, P., Chaudhary, K., Gautam, A., Kumar, R., Open Source Drug Discovery, C. and Raghava, G.P. (2013) In silico approach for predicting toxicity of peptides and proteins. *PLoS One*, **8**, e73957. doi: 10.1371/journal.pone.0073957.

39. Hasan, M., Islam, M. S., Chakraborty, S., Mustafa, A. H., Azim, K. F., Joy, Z. F., … Hasan, M. N. (2020). Contriving a chimeric polyvalent vaccine to prevent infections caused by Herpes Simplex Virus (Type-1 and Type-2): an exploratory

immunoinformatic approach. *Journal of Biomolecular Structure* and *Dynamics*, **38**(10):2898-2915. DOI: 10.1080/07391102.2019.1647286.

40. Hess, B., Bekker, H., Berendsen, H.J.C. and Fraaije, J.G.E.M. (1997) LINCS: A linear constraint solver for molecular simulations. *Journal* of *Computational Chemistry*, **18**, 1463-1472.doi:10.1002/(SICI)1096-987X(199709)18:12<1463::AID JCC4>3.0.CO;2-H

41. Hsu, H.S., Lin, T.H., Wu, H.Y., Lin, L.S., Chung, C.S., Chiou, M.T. and Lin, C.N., 2016. High detection rate of dog circovirus in diarrheal dogs. *BMC Veterinary Research*, *12*(1), p.116. DOI 10.1186/s12917-016-0722-8.

42. Humphrey, W., Dalke, A. and Schulten, K. (1996) VMD: Visual molecular dynamics. *Journal* of *Molecular Graphics and Modelling* , **14**, 33-38. doi: 10.1016/0263-7855(96)00018-5.

43. Ilyina, T.V. and Koonin, E.V., 1992. Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse replicons from eubacteria, eucaryotes and archaebacteria. *Nucleic acids research*, *20*(13), pp.3279-3285. DOI: 10.1093/nar/20.13.3279

44. Jacchieri, S.G., Torquato, R. and Brentani, R.R. (2003) Structural study of binding of flagellin by Toll-like receptor 5. *Journal of Bacteriology*, **185**, 4243-4247. doi: 10.1128/JB.185.14.4243-4247.2003.

45. Jain, P., Joshi, A., Akhtar, N., Krishnan, S. and Kaushik, V. (2021) An immunoinformatics study: designing multivalent T-cell epitope vaccine against Canine circovirus. *Journal of Genetic Engineering and Biotechnology*, **19**, 121. doi: 10.1186/s43141-021-00220-4.

46. Jenkins, G.M. and Holmes, E.C., 2003. The extent of codon usage bias in human RNA viruses and its evolutionary origin. *Virus research*, *92*(1), pp.1-7. DOI:10.1016/S0168- 1702(02)00309-X.

47. Jiang, M., Guo, J., Zhang, G., Jin, Q., Liu, Y., Jia, R., & Wang, A. (2020). Fine mapping of linear B cell epitopes on capsid protein of porcine circovirus 3. *Applied Microbiology and Biotechnology*, 104(14), 6223–6234. doi:10.1007/s00253-020-10664-2

48. Joshi, A., Pathak, D. C., Mannan, M. U., & Kaushik, V. (2021). In-silico designing of epitope-based vaccine against the seven banded grouper nervous necrosis virus affecting fish species. *Network Modeling Analysis in Health Informatics and Bioinformatics*, *10*(1), 1-12. doi: 10.1007/s13721-021-00315-5.

49. Joung, I.S. and Cheatham, T.E. (2008) Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *Journal of Physical Chemistry B*, **112**, 9020-9041. doi: 10.1021/jp8001614.

50. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Zidek, A., Potapenko, A. *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature*, **596**, 583-589. doi: 10.1038/s41586-021-03819-2.

51. Kalra, K., Gorle, S., Cavallo, L., Oliva, R. and Chawla, M. (2020) Occurrence and stability of lone pair-pi and OH-pi interactions between water and nucleobases in functional RNAs. *Nucleic Acids Research*, **48**, 5825-5838. doi: 10.1093/nar/gkaa345.

52. Kapoor, A., Dubovi, E. J., Henriquez-Rivera, J. A., & Lipkin, W. I. (2012). Complete Genome Sequence of the First Canine circovirus. *Journal of Virology*, 86(12), 7018–7018. doi:10.1128/jvi.00791-12

53. Karplus, M. and Kuriyan, J. (2005) Molecular dynamics and protein function. *Proceedings of the National Academy* of *Sciences of the United States of America*, **102**, 6679-6685.DOI: 10.1073/pnas.0408930102.

54. Karumathil, S., Raveendran, N.T., Ganesh, D., Kumar NS, S., Nair, R.R. and Dirisala, V.R., 2018. Evolution of Synonymous Codon Usage Bias in West African and Central African Strains of Monkeypox Virus. *Evolutionary Bioinformatics*, *14*, p.1176934318761368. DOI: 10.1177/1176934318761368.

55. Khandia, R., Singhal, S., Kumar, U., Ansari, A.I., Tiwari, R., Dhama, K., Das, J., Munjal, A. and Singh, R.K., 2019. Analysis of Nipah virus codon usage and adaptation to hosts. *Frontiers in microbiology*, *10*, p.886. DOI:10.3389/fmicb.2019.00886.

56. Kotsias, F., Bucafusco, D., Nuñez, D. A., Lago Borisovsky, L. A., Rodriguez, M., & Bratanich, A. C. (2019). Genomic characterization of Canine circovirus

associated with fatal disease in dogs in South America. *PloS one*, 14(6), e0218735. doi:10.1371/journal.pone.0218735

57. Krasovec, M. and Filatov, D.A., 2019. Evolution of Codon Usage Bias in Diatoms. *Genes*, *10*(11), p.894. DOI:10.3390/genes10110894.

58. Kumar, C.S. and Kumar, S., 2017. Synonymous codon usage of genes in polymerase complex of Newcastle disease virus. *Journal of basic microbiology*, *57*(6), pp.481-503. DOI:10.1002/jobm.201600740.

59. Kumar, N., Bera, B.C., Greenbaum, B.D., Bhatia, S., Sood, R., Selvaraj, P., Anand, T., Tripathi, B.N. and Virmani, N., 2016. Revelation of influencing factors in overall codon usage bias of equine influenza viruses. *PloS one*, *11*(4). DOI:10.1371/journal.pone.0154376.

60. Lamiable, A., Thévenet, P., Rey, J., Vavrusa, M., Derreumaux, P., & Tufféry, P. (2016). PEP-FOLD3: fasterde novostructure prediction for linear peptides in solution and in complex. *Nucleic Acids Research*, 44(W1), W449–W454. doi:10.1093/nar/gkw329

61. Lamolle, G., Fontenla, S., Rijo, G., Tort, J.F. and Smircich, P., 2019. Compositional analysis of flatworm genomes shows strong codon usage biases across all classes. *Frontiers in genetics*, *10*, p.771. DOI:10.3389/fgene.2019.00771.

62. Lee, J.Y., Han, G.G., Kim, E.B. and Choi, Y.J., 2017. Comparative genomics of Lactobacillus salivarius strains focusing on their host adaptation. *Microbiological research*, *205*, pp.48-58. DOI:10.1016/j.micres.2017.08.008.

63. Li, G., Wang, H., Wang, S., Xing, G., Zhang, C., Zhang, W., Liu, J., Zhang, J., Su, S. and Zhou, J., 2018. Insights into the genetic and host adaptability of emerging porcine circovirus 3. *Virulence*, *9*(1), pp.1301-1313. DOI:10.1080/21505594.2018.1492863.

64. Li, G., Zhang, W., Wang, R., Xing, G., Wang, S., Ji, X., ... & Zhou, J. (2019). Genetic analysis and evolutionary changes of the torque teno sus virus. *International journal of molecular sciences*, *20*(12), 2881. doi: 10.3390/ijms20122881.

65. Li, L., McGraw, S., Zhu, K., Leutenegger, C. M., Marks, S. L., Kubiski, S., … Pesavento, P. A. (2013). Circovirus in Tissues of Dogs with Vasculitis and

Hemorrhage. *Emerging Infectious Diseases*, 19(4), 534–541. doi:10.3201/eid1904.121390

66. Liao, P.C., Wang, K.K., Tsai, S.S., Liu, H.J., Huang, B.H. and Chuang, K.P., 2015. Recurrent positive selection and heterogeneous codon usage bias events leading to coexistence of divergent pigeon circoviruses. *Journal of General Virology*, *96*(8), pp.2262- 2273. DOI:10.1099/vir.0.000163.

67. Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J.L., Dror, R.O. and Shaw, D.E. (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins*, **78**, 1950-1958. doi: 10.1002/prot.22711.

68. Lister, R., Pelizzola, M., Dowen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.M. and Edsall, L., 2009. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, *462*(7271), pp.315-322.DOI: 10.1038/nature08514.

69. Liu, X.S., Zhang, Y.G., Fang, Y.Z. and Wang, Y.L., 2012. Patterns and influencing factor of synonymous codon usage in porcine circovirus. *Virology journal*, *9*(1), p.68. DOI:10.1186/1743-422X-9-68.

70. Liu, Y.S., Zhou, J.H., Chen, H.T., Ma, L.N., Pejsak, Z., Ding, Y.Z. and Zhang, J., 2011. The characteristics of the synonymous codon usage in enterovirus 71 virus and the effects of host on the virus in codon usage pattern. *Infection, Genetics and Evolution*, *11*(5), pp.1168-1173. DOI: 10.1016/j.meegid.2011.02.018.

71. Ma, J.J., Zhao, F., Zhang, J., Zhou, J.H., Ma, L.N., Ding, Y.Z., Chen, H.T., Gu, Y.X. and Liu, Y.S., 2013. Analysis of synonymous codon usage in dengue viruses. *Journal* of *Animal Veterinary advances*, *12*, pp.88-98. DOI: 10.3923/javaa.2013.88.98.

72. Magnan, C. N., Randall, A., & Baldi, P. (2009). SOLpro: accurate sequence-based prediction of protein solubility. *Bioinformatics*, *25*(17), 2200-2207. doi: 10.1093/bioinformatics/btp386.

73. Martin, A., Bertranpetit, J., Oliver, J.L. and Medina, J.R., 1989. Variation in G+ C- content and codon choice: differences among synonymous codon groups in vertebrate genes. *Nucleic acids research*, *17*(15), pp.6181-6189. DOI: 10.1093/nar/17.15.6181.

74. Moratorio, G., Iriarte, A., Moreno, P., Musto, H. and Cristina, J., 2013. A detailed comparative analysis on the overall codon usage patterns in West Nile virus. *Infection, Genetics and Evolution*, *14*, pp.396-400. DOI: 10.1016/j.meegid.2013.01.001.

75. Nasrullah, I., Butt, A.M., Tahir, S., Idrees, M. and Tong, Y., 2015. Genomic analysis of codon usage shows influence of mutation pressure, natural selection, and host features on Marburg virus evolution. *BMC evolutionary biology*, *15*(1), p.174. DOI:10.1186/s12862-015-0456-4.

76. Negahdaripour, M., Eslami, M., Nezafat, N., Hajighahramani, N., Ghoshoon, M. B., Shoolian, E., … Ghasemi, Y. (2017). A novel HPV prophylactic peptide vaccine, designed by immunoinformatics and structural vaccinology approaches. *Infection, Genetics and Evolution*, 54, 402–416. doi:10.1016/j.meegid.2017.08.002

77. Niagro Niagro, F.D., Forsthoefel, A.N., Lawther, R.P., Kamalanathan, L., Ritchie, B.W., Latimer, K.S. and Lukert, P.D., 1998. Beak and feather disease virus and porcine circovirus genomes: intermediates between the geminiviruses and plant circoviruses. *Archives of virology*, *143*(9), pp.1723-1744. DOI:10.1007/s007050050412.

78. Oliva, R., Shaikh, A.R., Petta, A., Vangone, A. and Cavallo, L. (2021) D936Y and Other Mutations in the Fusion Core of the SARS-CoV-2 Spike Protein Heptad Repeat 1: Frequency, Geographical Distribution, and Structural Effect. *Molecules*, **26**. doi: 10.3390/molecules26092622.

79. Oyarzún, P., & Kobe, B. (2016). Recombinant and epitope-based vaccines on the road to the market and implications for vaccine design and production. *Human vaccines & immunotherapeutics*, *12*(3), 763-767. doi: 10.1080/21645515.2015.1094595.

80. Parrinello, M. and Rahman, A. (1981) Polymorphic Transitions in Single-Crystals - a New Molecular-Dynamics Method. *Journal of Applied Physics* , **52**, 7182-7190.DOI: 10.1063/1.328693.

81. Piewbang, C., Jo, W.K., Puff, C., van der Vries, E., Kesdangsakonwut, S., Rungsipipat, A., Kruppa, J., Jung, K., Baumgärtner, W., Techangamsuwan, S. and

Ludlow, M., 2018. Novel Canine circovirus strains from Thailand: Evidence for genetic recombination. *Scientific reports*, *8*(1), pp.1-9. DOI:10.1038/s41598-018-25936-1.

82. Plog, S., Kubiski, S., Enders, J., Lübke-Becker, A., Pesavento, P.A. and Gruber, A.D., 2017. Detection of Dog Circovirus in a Dog Without Typical Lesions–What Do We Know About Subclinical Infection?. *Journal of Comparative Pathology*, *1*(156), p.75. DOI:10.1016/j.jcpa.2016.11.058.

83. Reynisson, B., Alvarez, B., Paul, S., Peters, B., & Nielsen, M. (2020). NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Research*, 48(W1), W449–W454. doi:10.1093/nar/gkaa379.

84. Reynisson, B., Barra, C., Kaabinejadian, S., Hildebrand, W.H., Peters, B. and Nielsen, M. (2020) Improved Prediction of MHC II Antigen Presentation through Integration and Motif Deconvolution of Mass Spectrometry MHC Eluted Ligand Data *Journal* of *Proteome Research*, **19**, 2304-2315. doi: 10.1021/acs.jproteome.9b00874.

85. Roy, A., Banerjee, R. and Basak, S., 2017. HIV progression depends on codon and amino acid usage profile of envelope protein and associated host-genetic influence. *Frontiersin microbiology*, *8*, p.1083. DOI:10.3389/fmicb.2017.01083.

86. Roy, A., Mukhopadhyay, S., Sarkar, I. and Sen, A., 2015. Comparative investigation of the various determinants that influence the codon and amino acid usage patterns in the genus Bifidobacterium. *World Journal of Microbiology and Biotechnology*, *31*(6), pp.959- 981. DOI:10.1007/s11274- 015-1850-1.

87. Schneidman-Duhovny, D., Inbar, Y., Nussinov, R., & Wolfson, H. J. (2005). PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Research*, 33(Web Server), W363–W367. doi:10.1093/nar/gki481

88. Segalés, J., 2012. Porcine circovirus type 2 (PCV2) infections: clinical signs, pathology and laboratory diagnosis. *Virus research*, *164*(1-2), pp.10-19. DOI:10.1016/j.virusres.2011.10.007.

89. Shackelton, L.A., Parrish, C.R. and Holmes, E.C., 2006. Evolutionary basis of codon usage and nucleotide composition bias in vertebrate DNA viruses. *Journal of molecular evolution*, *62*(5), pp.551-563. DOI:10.1007/s00239-005-0221-1.

90. Sharp, P.M., Cowe, E., Higgins, D.G., Shields, D.C., Wolfe, K.H. and Wright, F., 1988. Codon usage patterns in Escherichia coli, Bacillus subtilis, Saccharomyces cerevisiae, Schizosaccharomyces pombe, Drosophila melanogaster and Homo sapiens; a review of the considerable within-species diversity. *Nucleic acids research*, *16*(17), pp.8207-8211. DOI:10.1093/nar/16.17.8207.

91. Sharp, P.M., Stenico, M., Peden, J.F. and Lloyd, A.T., 1993. Codon usage: mutationalbias, translational selection, or both?. DOI:10.1042/bst0210835.

92. Song, T., Hao, J., Zhang, R., Tang, M., Li, W., Hui, W., Fu, Q., Wang, C., Xin, S., Zhang, S. and Rui, P., 2019. First detection and phylogenetic analysis of porcine circovirus type 2 in raccoon dogs. *BMC veterinary research*, *15*(1), p.107. DOI:10.1186/s12917-019- 18562.

93. Soria-Guerra, R.E., Nieto-Gomez, R., Govea-Alonso, D.O. and Rosales-Mendoza, S., 2015. An overview of bioinformatics tools for epitope prediction: implications on vaccine development. *Journal of biomedical informatics*, *53*, pp.405-414. DOI: 10.1016/j.jbi.2014.11.003.

94. Sun, W., Wang, W., Cao, L., Zheng, M., Zhuang, X., Zhang, H., … Jin, N. (2020). Genetic characterization of three porcine circovirus-like viruses in pigs with diarrhoea in China. *Transboundary and Emerging Diseases*. doi:10.1111/tbed.13731

95. Sun, W., Zhang, H., Zheng, M., Cao, H., Lu, H., Zhao, G., Xie, C., Cao, L., Wei, X., Bi, J. and Yi, C., 2019. The detection of Canine circovirus in Guangxi, China. *Virus research*, *259*, pp.85-89. DOI:10.1016/j.virusres.2018.10.021.

96. Tao, P., Dai, L., Luo, M., Tang, F., Tien, P. and Pan, Z., 2009. Analysis of synonymous codon usage in classical swine fever virus. *Virus genes*, *38*(1), pp.104-112. DOI: 10.1007/s11262-008-0296-z.

97. Thaiwong, T., Wise, A.G., Maes, R.K., Mullaney, T. and Kiupel, M., 2016. Canine circovirus 1 (CaCV-1) and Canine Parvovirus 2 (CPV-2) Recurrent Dual Infections in a Papillon Breeding Colony. *Veterinary pathology*, *53*(6), pp.1204-1209. DOI:10.1177/0300985816646430.

98. Todd, D., 2000. Circoviruses: immunosuppressive threats to avian species: a review. *Avian Pathology*, *29*(5), pp.373-394. DOI:10.1080/030794500750047126.

99. Todd, D., McNulty, M., Adair, B., & Allan, G. . (2001). Animal circoviruses. *Advances in Virus Research*, 1–70. doi:10.1016/s0065-3527(01)57000-1

100. Tree, M.O., McKellar, D.R., Kieft, K.J., Watson, A.M., Ryman, K.D. and Conway, M.J. (2016) Insect-specific flavivirus infection is restricted by innate immunity in the vertebrate host. *Virology*, **497**, 81-91. doi: 10.1016/j.virol.2016.07.005.

101. Trott, O., & Olson, A. J. (2009). AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, NA–NA. doi:10.1002/jcc.21334

102. Tsai, G.T., Lin, Y.C., Lin, W.H., Lin, J.H., Chiou, M.T., Liu, H.F. and Lin, C.N., 2019. Phylogeographic and genetic characterization of porcine circovirus type 2 in Taiwan from 2001–2017. *Scientific reports*, *9*(1), pp.1-13. DOI:10.1038/s41598-019-47209-1.

103. Van Zundert, G.C.P., Rodrigues, J., Trellet, M., Schmitz, C., Kastritis, P.L., Karaca, E., Melquiond, A.S.J., van Dijk, M., de Vries, S.J. and Bonvin, A. (2016) The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *Journal of Molecular Biology*, **428**, 720-725. doi: 10.1016/j.jmb.2015.09.014.

104. Vangone, A., Oliva, R. and Cavallo, L. (2012) CONS-COCOMAPS: a novel tool to measure and visualize the conservation of inter-residue contacts in multiple docking solutions. *Bmc Bioinformatics*, **13**. DOI: 10.1186/1471-2105-13-S4-S19.

105. Vangone, A., Spinelli, R., Scarano, V., Cavallo, L. and Oliva, R. (2011) COCOMAPS: a web application to analyze and visualize contacts at the interface of biomolecular complexes. *Bioinformatics*, **27**, 2915-2916. doi: 10.1093/bioinformatics/btr484.

106. Xu, X.Z., Liu, Q.P., Fan, L.J., Cui, X.F. and Zhou, X.P., 2008. Analysis of synonymous codon usage and evolution of begomoviruses. *Journal of Zhejiang University Science B*, *9*(9), pp.667-674. DOI: 10.1631/jzus.B0820005.

107. Xu, Y., Jia, R., Zhang, Z., Lu, Y., Wang, M., Zhu, D., Chen, S., Liu, M., Yin, Z. and Cheng, A., 2015. Analysis of synonymous codon usage pattern in duck circovirus. *Gene*, *557*(2), pp.138-145. DOI:10.1016/j.gene.2014.12.019.

108. Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., & Zhang, Y. (2014). The I-TASSER Suite: protein structure and function prediction. *Nature Methods*, 12(1), 7–8. doi:10.1038/nmeth.3213

109. Yao, H., Chen, M. and Tang, Z., 2019. Analysis of Synonymous Codon Usage Bias in Flaviviridae Virus. *BioMed research international*. DOI:10.1155/2019/5857285.

110. Yazdani, Z., Rafiei, A., Valadan, R., Ashrafi, H., Pasandi, M., & Kardan, M. (2020). Designing a potent L1 protein-based HPV peptide vaccine: A bioinformatics approach. *Computational Biology and Chemistry*, 85, 107209. doi:10.1016/j.compbiolchem.2020.107209

111. Zaccaria, G., Malatesta, D., Scipioni, G., Di Felice, E., Campolo, M., Casaccia, C., Savini, G., Di Sabatino, D. and Lorusso, A., 2016. Circovirus in domestic and wild carnivores: an important opportunistic agent?. *Virology*, *490*, pp.69-74. DOI: 10.1016/j.virol.2016.01.007.

112. Zhang, J., Wang, M., Liu, W.Q., Zhou, J.H., Chen, H.T., Ma, L.N., Ding, Y.Z., Gu, Y.X. and Liu, Y.S., 2011. Analysis of codon usage and nucleotide composition bias in polioviruses. *Virology journal*, *8*(1), p.146. DOI: 10.1186/1743-422X-8-146.

Supplementary table *S1: Nucleotides composition of CanineCV genome at third codon position (T3%, G3%, C3% and A3%), overall AT% and GC% composition along with GC12 (the average of GC1 and GC2) and GC3% analysis.

Coding Sequences (CDSs)/ Accession number CanineCV	Gene	Length	A%	C%	T%	G%	A3%	C3%	T3%	G3%	AT%	GC%	GC1	GC2	GC12	AT3%	GC3%	Nc	CAI	GRAVY	AROMO
Coding sequence C																					
JQ821392	C		32.96	24.85	21.40	20.79	38.01	20.30	22.51	19.19	51.66	45.63	48.34	49.08	48.71	60.52	39.48	49.70	0.50	0.98	0.111111
KC241982			31.98	25.22	21.89	20.91	36.53	21.03	23.62	18.82	50.92	46.13	49.08	49.45	49.26	60.15	39.85	48.20	0.51	0.97	0.111111
KC241983			31.12	25.95	21.53	21.40	33.95	22.88	22.51	20.66	50.92	47.36	49.08	49.45	49.26	56.46	43.54	54.30	0.50	0.97	0.118519
KC241984			32.10	24.97	21.40	21.53	36.16	21.40	22.14	20.30	51.29	46.49	48.71	49.08	48.89	58.30	41.70	50.80	0.50	0.97	0.111111
NC_020904			31.98	25.22	21.89	20.91	36.53	21.03	23.62	18.82	50.92	46.13	49.08	49.45	49.26	60.15	39.85	48.20	0.51	0.97	0.111111
MF457592			32.60	25.34	21.03	21.03	37.64	21.03	21.40	19.93	50.92	46.37	49.08	49.08	49.08	59.04	40.96	47.80	0.51	0.99	0.111111
MK033608			31.98	24.85	21.89	21.28	35.79	20.30	23.62	20.30	51.29	46.13	48.71	49.08	48.89	59.41	40.59	50.60	0.50	0.96	0.111111
KJ530972			32.23	25.09	21.28	21.40	36.53	20.66	22.14	20.66	51.66	46.49	48.34	49.82	49.08	58.67	41.33	51.40	0.50	0.97	0.111111
KT734812			33.33	24.97	21.16	20.54	39.48	21.03	21.40	18.08	52.03	45.51	47.97	49.45	48.71	60.89	39.11	48.40	0.50	0.99	0.111111
KT734813			32.47	24.97	21.77	20.79	37.27	20.30	22.88	19.56	51.66	45.76	48.34	49.08	48.71	60.15	39.85	49.20	0.50	0.97	0.111111
KT734814			32.47	24.97	21.28	21.28	37.64	20.66	21.77	19.93	51.66	46.25	48.34	49.82	49.08	59.41	40.59	50.00	0.50	0.97	0.111111
KT734815			32.35	25.09	21.16	21.40	36.90	21.03	21.40	20.66	51.29	46.49	48.71	49.08	48.89	58.30	41.70	51.60	0.51	0.98	0.111111
KT734816			32.23	24.97	21.65	21.16	35.79	21.03	22.51	20.66	52.03	46.13	47.97	48.71	48.34	58.30	41.70	51.60	0.50	0.97	0.111111
KT734817			32.60	24.85	21.28	21.28	37.64	20.30	21.77	20.30	51.66	46.13	48.34	49.45	48.89	59.41	40.59	50.00	0.49	0.97	0.111111
KT734818			32.60	24.97	21.16	21.28	37.64	20.66	21.40	20.30	51.66	46.25	48.34	49.45	48.89	59.04	40.96	50.00	0.50	0.98	0.111111
KT734819			32.72	25.09	21.16	21.03	36.53	21.03	21.77	20.66	52.40	46.13	47.60	49.08	48.34	58.30	41.70	52.40	0.50	0.98	0.111111
KT734820			32.35	25.09	20.91	21.65	36.90	21.03	20.66	21.40	51.66	46.74	48.34	49.45	48.89	57.56	42.44	51.60	0.50	0.98	0.111111
KT734821			32.60	24.97	21.40	21.03	36.90	20.66	21.77	20.66	51.66	46.00	48.34	48.34	48.34	58.67	41.33	49.90	0.51	0.98	0.111111
KT734822			32.10	25.34	21.28	21.28	35.79	21.03	21.77	21.40	50.92	46.62	49.08	48.34	48.71	57.56	42.44	53.10	0.50	0.98	0.111111
KT734823			33.09	24.97	21.28	20.66	39.11	20.66	22.14	18.08	51.29	45.63	48.71	49.45	49.08	61.25	38.75	48.60	0.50	0.99	0.107407
KT734824			32.35	25.09	20.91	21.65	36.90	21.03	20.66	21.40	51.66	46.74	48.34	49.45	48.89	57.56	42.44	51.60	0.50	0.98	0.111111
KT734825			32.60	24.97	21.40	21.03	36.90	20.66	21.77	20.66	51.66	46.00	48.34	48.34	48.34	58.67	41.33	49.90	0.51	0.98	0.111111
KT734826			33.09	24.97	21.28	20.66	39.11	20.66	22.14	18.08	51.29	45.63	48.71	49.45	49.08	61.25	38.75	48.60	0.50	0.99	0.107407

KT734827			32.35	25.09	21.16	21.40	36.90	21.03	21.40	20.66	51.29	46.49	48.71	49.08	48.89	58.30	41.70	51.60	0.51	0.98	0.111111
KT734828			33.33	24.97	21.16	20.54	39.48	21.03	21.40	18.08	52.03	45.51	47.97	49.45	48.71	60.89	39.11	48.40	0.50	0.99	0.111111
MH454599			31.61	25.83	20.66	21.89	35.06	21.77	21.03	22.14	49.08	47.72	50.92	48.34	49.63	56.09	43.91	51.30	0.50	0.98	0.111111
KY388483			31.61	25.09	21.16	22.14	33.21	21.77	21.03	23.99	51.29	47.23	48.71	47.23	47.97	54.24	45.76	54.20	0.52	0.96	0.111111
KT946839			32.35	25.58	21.03	21.03	35.42	21.40	22.14	21.03	50.92	46.62	49.08	48.34	48.71	57.56	42.44	52.00	0.53	0.99	0.111111
KY388484			31.49	26.08	20.79	21.65	34.32	23.99	20.66	21.03	50.18	47.72	49.82	48.34	49.08	54.98	45.02	53.00	0.51	0.99	0.114815
KY388485			31.37	25.95	20.79	21.89	33.95	23.99	20.66	21.40	50.18	47.85	49.82	48.34	49.08	54.61	45.39	52.90	0.52	0.98	0.114815
KY388487			30.38	25.83	21.28	22.51	31.37	23.25	21.77	23.62	50.18	48.34	49.82	48.34	49.08	53.14	46.86	52.40	0.52	0.95	0.118519
KY388491			31.86	25.09	21.03	22.02	33.58	21.77	21.03	23.62	51.29	47.11	48.71	47.23	47.97	54.61	45.39	52.90	0.52	0.97	0.111111
KY388492			31.73	25.22	21.03	22.02	33.58	21.77	21.03	23.62	50.92	47.23	49.08	47.23	48.15	54.61	45.39	55.00	0.51	0.97	0.111111
KY388493			31.73	25.09	21.03	22.14	33.21	21.77	21.03	23.99	51.29	47.23	48.71	47.23	47.97	54.24	45.76	53.80	0.52	0.96	0.111111
KY388497			31.73	25.22	21.03	22.02	33.58	21.77	21.03	23.62	50.92	47.23	49.08	47.23	48.15	54.61	45.39	55.00	0.51	0.97	0.111111
KY388499			31.98	24.85	21.16	22.02	33.58	21.40	21.03	23.99	52.03	46.86	47.97	47.23	47.60	54.61	45.39	53.60	0.52	0.96	0.111111
KY388500			31.37	26.08	20.66	21.89	33.58	24.35	20.66	21.40	50.18	47.97	49.82	48.34	49.08	54.24	45.76	52.20	0.52	0.98	0.114815
KY388501			31.37	26.20	20.66	21.77	33.95	24.35	20.30	21.40	50.18	47.97	49.82	48.34	49.08	54.24	45.76	53.60	0.52	0.99	0.114815
KY388502			31.73	25.83	20.79	21.65	34.69	23.99	20.66	20.66	50.18	47.48	49.82	47.97	48.89	55.35	44.65	52.60	0.51	0.98	0.114815
KY388503			31.86	25.09	21.03	22.02	33.21	21.77	21.03	23.99	51.66	47.11	48.34	47.23	47.79	54.24	45.76	54.00	0.52	0.97	0.111111
KY388480			31.49	25.46	21.77	21.28	34.32	21.40	22.88	21.40	51.66	46.74	48.34	49.08	48.71	57.20	42.80	53.00	0.52	0.97	0.111111
KY388481			32.35	25.83	20.79	21.03	35.06	21.77	21.77	21.40	50.92	46.86	49.08	48.34	48.71	56.83	43.17	52.60	0.53	1.00	0.107407
KY388482			32.23	25.22	21.28	21.28	35.42	20.66	22.51	21.40	50.92	46.49	49.08	48.34	48.71	57.93	42.07	51.40	0.52	0.98	0.111111
KY388486			31.24	25.95	21.03	21.77	33.95	24.72	20.30	21.03	51.66	47.72	48.34	49.08	48.71	54.24	45.76	52.80	0.54	0.98	0.114815
KY388488			32.35	25.34	20.30	22.02	35.06	23.25	18.82	22.88	50.92	47.36	49.08	46.86	47.97	53.87	46.13	54.80	0.52	0.99	0.111111
KY388489			32.23	25.46	21.16	21.16	35.06	21.03	22.51	21.40	50.92	46.62	49.08	48.34	48.71	57.56	42.44	52.30	0.53	0.98	0.111111
KY388490			32.23	25.83	20.79	21.16	35.06	21.77	21.77	21.40	50.55	46.99	49.45	48.34	48.89	56.83	43.17	52.60	0.53	1.00	0.107407
KY388494			31.12	25.58	20.91	22.39	33.21	21.77	21.77	23.25	49.08	47.97	50.92	47.97	49.45	54.98	45.02	51.10	0.51	0.96	0.111111
KY388495			30.87	26.08	20.79	22.26	32.84	22.14	21.03	23.99	49.45	48.34	50.55	48.34	49.45	53.87	46.13	50.30	0.51	0.97	0.111111
KY388496			30.87	26.08	20.79	22.26	32.84	22.14	21.03	23.99	49.45	48.34	50.55	48.34	49.45	53.87	46.13	50.30	0.51	0.97	0.111111
KY388498			31.37	26.08	20.54	22.02	33.95	24.35	19.93	21.77	50.18	48.09	49.82	48.34	49.08	53.87	46.13	53.10	0.52	0.98	0.114815
MG279132			32.60	25.58	20.66	21.16	36.53	22.88	20.30	20.30	50.92	46.74	49.08	47.97	48.52	56.83	43.17	51.20	0.51	1.00	0.114815
MG279133			32.60	25.58	20.66	21.16	36.53	22.88	20.30	20.30	50.92	46.74	49.08	47.97	48.52	56.83	43.17	51.50	0.51	1.00	0.114815
MG279134			32.60	25.58	20.66	21.16	36.53	22.88	20.30	20.30	50.92	46.74	49.08	47.97	48.52	56.83	43.17	51.20	0.51	1.00	0.114815

MG279136			32.47	25.58	20.54	21.40	36.53	22.88	20.30	20.30	50.55	46.99	49.45	48.34	48.89	56.83	43.17	51.20	0.51	0.99	0.114815
MG279137			32.35	25.58	20.66	21.40	36.53	22.88	20.30	20.30	50.55	46.99	49.45	48.34	48.89	56.83	43.17	51.50	0.51	0.99	0.111111
MG279138			32.60	25.71	20.54	21.16	36.53	23.25	19.93	20.30	50.92	46.86	49.08	47.97	48.52	56.46	43.54	51.40	0.51	1.00	0.114815
MG279139			32.84	25.58	20.66	20.91	36.90	22.88	20.30	19.93	50.92	46.49	49.08	47.60	48.34	57.20	42.80	51.10	0.51	1.00	0.114815
MG266899			32.23	25.09	21.28	21.40	36.16	21.40	21.40	21.03	52.40	46.49	47.60	49.45	48.52	57.56	42.44	52.50	0.51	0.97	0.111111
MF797786			31.98	25.09	21.28	21.65	34.69	20.66	22.51	22.14	50.92	46.74	49.08	48.34	48.71	57.20	42.80	50.30	0.51	0.97	0.111111
MG279118			31.98	25.83	20.05	22.14	35.06	23.99	18.08	22.88	50.92	47.97	49.08	47.97	48.52	53.14	46.86	52.80	0.52	0.99	0.111111
MG279119			32.10	25.58	21.03	21.28	36.16	22.88	21.03	19.93	50.55	46.86	49.45	48.34	48.89	57.20	42.80	53.30	0.50	0.99	0.114815
MG279120			32.10	25.34	21.28	21.28	36.16	22.51	21.40	19.93	50.92	46.62	49.08	48.34	48.71	57.56	42.44	53.80	0.50	0.98	0.114815
MG279121			32.10	25.58	21.03	21.28	36.16	22.88	21.03	19.93	50.55	46.86	49.45	48.34	48.89	57.20	42.80	53.30	0.50	0.99	0.114815
MG279122			31.98	25.58	21.16	21.28	36.16	22.88	21.03	19.93	50.55	46.86	49.45	48.34	48.89	57.20	42.80	53.70	0.50	0.98	0.114815
MG279123			31.98	25.58	21.16	21.28	36.16	22.88	21.03	19.93	50.55	46.86	49.45	48.34	48.89	57.20	42.80	53.70	0.50	0.98	0.114815
MG279125			32.10	25.58	21.03	21.28	36.16	22.88	21.03	19.93	50.55	46.86	49.45	48.34	48.89	57.20	42.80	53.30	0.50	0.99	0.114815
MG279126			32.35	25.58	20.91	21.16	36.16	22.88	21.03	19.93	50.92	46.74	49.08	48.34	48.71	57.20	42.80	53.50	0.50	0.99	0.114815
MG279127			32.10	25.58	21.03	21.28	36.16	22.88	21.03	19.93	50.55	46.86	49.45	48.34	48.89	57.20	42.80	53.30	0.50	0.99	
MG279128			32.23	25.58	20.91	21.28	36.53	22.88	20.66	19.93	50.55	46.86	49.45	48.34	48.89	57.20	42.80	53.10	0.50	0.99	
MG279129			32.10	25.58	21.03	21.28	36.16	22.88	21.03	19.93	50.55	46.86	49.45	48.34	48.89	57.20	42.80	53.30	0.50	0.99	
MG279130			31.98	25.58	21.16	21.28	36.16	22.88	21.03	19.93	50.55	46.86	49.45	48.34	48.89	57.20	42.80	53.70	0.50	0.98	
MG279131			32.10	25.58	21.03	21.28	36.16	22.88	21.03	19.93	50.55	46.86	49.45	48.34	48.89	57.20	42.80	53.30	0.50	0.99	
MG279135			32.23	25.46	21.03	21.28	36.53	22.51	21.03	19.93	50.55	46.74	49.45	48.34	48.89	57.56	42.44	53.10	0.50	0.98	
MG279140			31.98	25.58	21.03	21.40	36.16	22.88	21.03	19.93	50.55	46.99	49.45	48.71	49.08	57.20	42.80	53.80	0.50	0.98	
MG279141			31.98	25.71	20.91	21.40	36.16	22.88	20.66	20.30	50.55	47.11	49.45	48.71	49.08	56.83	43.17	54.40	0.50	0.99	
MK731981			31.00	26.32	21.28	21.40	34.32	23.25	21.40	21.03	50.18	47.72	49.82	49.08	49.45	55.72	44.28	52.00	0.51	0.98	
MK731982			30.63	26.08	20.79	22.51	31.73	22.51	21.40	24.35	49.45	48.59	50.55	48.34	49.45	53.14	46.86	52.00	0.51	0.97	
MN128702			31.00	25.83	21.16	22.02	33.58	21.77	22.14	22.51	49.45	47.85	50.55	48.71	49.63	55.72	44.28	50.90	0.50	0.97	
MK944079			31.49	25.83	21.28	21.40	34.32	21.40	22.88	21.40	49.45	47.23	50.55	48.34	49.45	57.20	42.80	51.50	0.50	0.98	
MK944080			31.00	26.45	21.16	21.40	34.69	23.62	21.03	20.66	50.18	47.85	49.82	49.45	49.63	55.72	44.28	52.30	0.51	0.99	
KP260925			31.00	25.58	20.66	22.76	32.47	22.51	20.66	24.35	49.82	48.34	50.18	47.97	49.08	53.14	46.86	53.20	0.52	0.96	

KP260926			31.12	25.46	20.54	22.88	32.84	21.77	21.03	24.35	49.82	48.34	50.18	48.71	49.45	53.87	46.13	54.20	0.51	0.96	
KP260927			31.12	25.46	20.54	22.88	32.47	21.77	21.40	24.35	49.82	48.34	50.18	48.71	49.45	53.87	46.13	54.50	0.51	0.96	
KF887949			32.35	25.22	21.40	21.03	35.79	21.40	22.14	20.66	52.03	46.25	47.97	48.71	48.34	57.93	42.07	50.60	0.50	0.98	
KT283604			31.98	24.35	22.02	21.65	36.16	19.93	23.25	20.66	51.29	46.00	48.71	48.71	48.71	59.41	40.59	49.20	0.50	0.94	
KP941114			31.12	26.08	20.17	22.63	31.73	23.99	20.66	23.62	48.34	48.71	51.66	46.86	49.26	52.40	47.60	54.40	0.52	0.98	
MK424788			33.09	24.97	21.03	20.91	38.75	20.30	21.40	19.56	51.29	45.88	48.71	49.08	48.89	60.15	39.85	48.50	0.51	0.99	
Coding sequence R																					
JQ821392	R		22.70	23.36	23.79	30.15	12.83	25.66	29.93	31.58	46.49	53.51	57.24	46.05	51.64	42.76	57.24	57.31	0.54	0.71	0.11
KC241982			23.25	23.36	23.57	29.82	13.49	25.99	29.28	31.25	46.82	53.18	56.91	45.39	51.15	42.76	57.24	55.28	0.55	0.72	0.11
KC241983			23.57	22.15	24.56	29.71	13.16	22.70	32.57	31.58	48.14	51.86	56.91	44.41	50.66	45.72	54.28	55.01	0.54	0.69	0.10
KC241984			23.68	23.46	23.03	29.82	13.82	26.97	28.29	30.92	46.71	53.29	56.91	45.07	50.99	42.11	57.89	56.56	0.55	0.73	0.11
NC_020904			23.25	23.36	23.57	29.82	13.49	25.99	29.28	31.25	46.82	53.18	56.91	45.39	51.15	42.76	57.24	55.28	0.55	0.72	0.11
MF457592			23.25	23.46	23.25	30.04	13.16	25.99	28.95	31.91	46.49	53.51	57.57	45.07	51.32	42.11	57.89	55.92	0.55	0.72	0.11
MK033608			24.01	23.25	23.57	29.17	14.47	25.99	29.28	30.26	47.59	52.41	56.25	44.74	50.49	43.75	56.25	55.13	0.55	0.73	0.11
KJ530972			23.25	23.14	23.68	29.93	13.49	26.32	29.28	30.92	46.93	53.07	56.58	45.39	50.99	42.76	57.24	55.83	0.55	0.71	0.11
KT734812			23.03	23.14	23.79	30.04	13.16	25.33	29.28	32.24	46.82	53.18	56.58	45.39	50.99	42.43	57.57	54.38	0.55	0.71	0.11
KT734813			23.68	23.46	23.03	29.82	13.49	26.97	27.30	32.24	46.71	53.29	56.25	44.41	50.33	40.79	59.21	54.24	0.57	0.73	0.11
KT734814			23.79	23.03	23.68	29.50	13.49	25.99	30.26	30.26	47.48	52.52	56.25	45.07	50.66	43.75	56.25	55.43	0.55	0.72	0.11
KT734815			23.57	22.70	23.79	29.93	14.14	25.33	29.93	30.59	47.37	52.63	56.91	45.07	50.99	44.08	55.92	55.09	0.55	0.71	0.11
KT734816			23.68	23.57	23.14	29.61	13.82	26.97	28.29	30.92	46.82	53.18	56.58	45.07	50.82	42.11	57.89	55.31	0.56	0.74	0.11
KT734817			23.68	23.14	23.79	29.39	13.49	26.97	29.28	30.26	47.48	52.52	55.59	44.74	50.16	42.76	57.24	54.49	0.55	0.72	0.11
KT734818			23.68	23.14	23.79	29.39	13.49	26.97	29.28	30.26	47.48	52.52	55.59	44.74	50.16	42.76	57.24	54.49	0.55	0.72	0.11
KT734819			23.46	23.36	23.57	29.61	13.82	27.30	28.29	30.59	47.04	52.96	55.59	45.39	50.49	42.11	57.89	54.73	0.55	0.72	0.11
KT734820			23.25	23.68	23.25	29.82	13.16	27.30	28.62	30.92	46.49	53.51	56.91	45.39	51.15	41.78	58.22	56.78	0.55	0.73	0.11
KT734821			23.46	22.70	24.12	29.71	13.82	25.00	29.93	31.25	47.59	52.41	56.58	44.41	50.49	43.75	56.25	57.92	0.56	0.70	0.11
KT734822			23.46	23.14	23.79	29.61	13.16	26.64	29.61	30.59	47.26	52.74	55.92	45.07	50.49	42.76	57.24	54.79	0.55	0.72	0.11
KT734823			23.03	23.03	23.90	30.04	13.49	25.00	29.28	32.24	46.93	53.07	56.58	45.39	50.99	42.76	57.24	54.50	0.54	0.70	0.11
KT734824			23.25	23.68	23.25	29.82	13.16	27.30	28.62	30.92	46.49	53.51	56.91	45.39	51.15	41.78	58.22	56.78	0.55	0.73	0.11
KT734825			23.46	22.70	24.12	29.71	13.82	25.00	29.93	31.25	47.59	52.41	56.58	44.41	50.49	43.75	56.25	57.92	0.56	0.70	0.11

KT734826			23.03	23.03	23.90	30.04	13.49	25.00	29.28	32.24	46.93	53.07	56.58	45.39	50.99	42.76	57.24	54.50	0.54	0.70	0.11
KT734827			23.57	22.70	23.79	29.93	14.14	25.33	29.93	30.59	47.37	52.63	56.91	45.07	50.99	44.08	55.92	55.09	0.55	0.71	0.11
KT734828			23.03	23.14	23.79	30.04	13.16	25.33	29.28	32.24	46.82	53.18	56.58	45.39	50.99	42.43	57.57	54.38	0.55	0.71	0.11
MH454599			23.68	22.37	24.67	29.28	13.49	25.00	31.91	29.61	48.36	51.64	58.22	42.11	50.16	45.39	54.61	59.40	0.54	0.70	0.11
KY388483			22.70	22.04	24.89	30.37	11.51	23.68	31.91	32.89	47.59	52.41	56.58	44.08	50.33	43.42	56.58	57.42	0.54	0.66	0.11
KT946839			23.36	22.92	23.68	30.04	12.83	25.33	29.93	31.91	47.04	52.96	56.91	44.74	50.82	42.76	57.24	56.36	0.56	0.71	0.11
KY388484			22.04	22.26	25.11	30.59	10.20	24.01	32.24	33.55	47.15	52.85	56.25	44.74	50.49	42.43	57.57	55.34	0.53	0.66	0.11
KY388485			24.12	22.81	23.57	29.50	14.47	25.33	29.28	30.92	47.70	52.30	57.24	43.42	50.33	43.75	56.25	58.17	0.55	0.72	0.11
KY388487			22.81	21.71	25.33	30.15	11.84	23.36	32.24	32.57	48.14	51.86	55.59	44.08	49.84	44.08	55.92	56.71	0.53	0.66	0.11
KY388491			22.59	22.15	25.00	30.26	11.18	23.68	32.24	32.89	47.59	52.41	56.58	44.08	50.33	43.42	56.58	57.66	0.54	0.66	0.11
KY388492			22.15	21.82	25.33	30.70	10.20	22.37	33.22	34.21	47.48	52.52	56.58	44.41	50.49	43.42	56.58	55.48	0.53	0.64	0.11
KY388493			22.70	21.93	24.89	30.48	11.51	23.36	31.91	33.22	47.59	52.41	56.58	44.08	50.33	43.42	56.58	57.39	0.54	0.66	0.11
KY388497			22.15	22.04	25.22	30.59	10.53	23.03	32.89	33.55	47.37	52.63	56.58	44.74	50.66	43.42	56.58	55.76	0.52	0.65	0.11
KY388499			22.70	22.15	24.89	30.26	11.18	23.68	32.24	32.89	47.59	52.41	56.58	44.08	50.33	43.42	56.58	57.40	0.54	0.67	0.11
KY388500			22.15	22.26	25.00	30.59	10.20	24.01	31.91	33.88	47.15	52.85	56.25	44.41	50.33	42.11	57.89	55.17	0.53	0.66	0.11
KY388501			22.70	22.48	24.56	30.26	11.51	24.34	31.25	32.89	47.26	52.74	56.91	44.08	50.49	42.76	57.24	58.71	0.53	0.68	0.11
KY388502			22.15	22.26	25.00	30.59	10.20	24.01	31.91	33.88	47.15	52.85	56.25	44.41	50.33	42.11	57.89	55.17	0.53	0.66	0.11
KY388503			22.70	22.26	24.78	30.26	11.51	24.01	31.58	32.89	47.48	52.52	56.58	44.08	50.33	43.09	56.91	58.02	0.54	0.67	0.11
KY388480			24.01	22.48	24.12	29.39	14.14	25.00	30.26	30.59	48.14	51.86	55.92	44.08	50.00	44.41	55.59	57.21	0.56	0.71	0.11
KY388481			23.36	22.70	23.90	30.04	12.83	25.00	30.26	31.91	47.26	52.74	56.91	44.41	50.66	43.09	56.91	56.30	0.56	0.70	0.11
KY388482			24.01	22.59	24.12	29.28	14.14	25.33	30.26	30.26	48.14	51.86	55.92	44.08	50.00	44.41	55.59	57.06	0.56	0.71	0.11
KY388486			22.92	22.70	24.67	29.71	12.17	25.33	31.25	31.25	47.59	52.41	56.58	44.08	50.33	43.42	56.58	57.61	0.53	0.69	0.11
KY388488			22.37	23.14	24.23	30.26	9.87	26.97	30.59	32.57	46.60	53.40	56.58	44.08	50.33	40.46	59.54	54.14	0.55	0.69	0.11
KY388489			23.25	22.59	23.90	30.26	12.83	24.67	30.26	32.24	47.15	52.85	56.91	44.74	50.82	43.09	56.91	56.37	0.56	0.69	0.11
KY388490			23.46	22.81	23.79	29.93	12.83	25.33	29.93	31.91	47.26	52.74	56.91	44.08	50.49	42.76	57.24	56.57	0.56	0.71	0.11
KY388494			23.57	23.36	23.14	29.93	13.16	26.97	28.62	31.25	46.71	53.29	58.22	43.42	50.82	41.78	58.22	57.12	0.55	0.73	0.11
KY388495			23.68	23.03	23.46	29.82	13.49	26.32	29.28	30.92	47.15	52.85	57.57	43.75	50.66	42.76	57.24	56.54	0.55	0.72	0.11
KY388496			23.57	23.03	23.46	29.93	13.16	26.32	29.28	31.25	47.04	52.96	57.57	43.75	50.66	42.43	57.57	56.45	0.55	0.72	0.11
KY388498			23.57	23.36	23.79	29.28	13.16	26.97	29.61	30.26	47.37	52.63	56.25	44.41	50.33	42.76	57.24	56.24	0.56	0.72	0.11
MG279132			22.81	22.92	24.01	30.26	12.83	24.67	29.61	32.89	46.82	53.18	56.91	45.07	50.99	42.43	57.57	59.37	0.53	0.69	0.11
MG279133			22.92	22.81	24.01	30.26	12.83	24.67	29.61	32.89	46.93	53.07	56.91	44.74	50.82	42.43	57.57	58.86	0.53	0.69	0.11

MG279134			22.92	22.81	24.01	30.26	12.83	24.67	29.61	32.89	46.93	53.07	56.91	44.74	50.82	42.43	57.57	58.86	0.53	0.69	0.11
MG279136			22.81	22.81	24.12	30.26	12.83	24.34	29.93	32.89	46.93	53.07	56.91	45.07	50.99	42.76	57.24	59.70	0.53	0.69	0.11
MG279137			22.92	22.81	24.01	30.26	12.83	24.67	29.61	32.89	46.93	53.07	56.91	44.74	50.82	42.43	57.57	58.86	0.53	0.69	0.11
MG279138			22.92	22.92	24.01	30.15	12.83	24.67	29.61	32.89	46.93	53.07	56.91	44.74	50.82	42.43	57.57	58.87	0.53	0.70	0.11
MG279139			23.14	22.70	24.01	30.15	13.16	24.67	29.61	32.57	47.15	52.85	56.58	44.74	50.66	42.76	57.24	59.08	0.53	0.70	0.11
MG266899			23.14	23.90	23.68	29.28	13.49	27.30	29.61	29.61	46.82	53.18	57.57	45.07	51.32	43.09	56.91	55.47	0.55	0.73	0.11
MF797786			23.90	22.26	24.01	29.82	13.82	24.01	30.59	31.58	47.92	52.08	56.25	44.41	50.33	44.41	55.59	57.45	0.55	0.70	0.11
MG279118			22.48	23.25	24.23	30.04	9.87	26.97	30.59	32.57	46.71	53.29	56.58	43.75	50.16	40.46	59.54	54.05	0.55	0.70	0.11
MG279119			22.37	22.81	24.45	30.37	10.20	25.66	31.25	32.89	46.82	53.18	56.25	44.74	50.49	41.45	58.55	55.86	0.55	0.68	0.11
MG279120			22.04	22.81	24.67	30.48	10.20	25.33	31.58	32.89	46.71	53.29	56.58	45.07	50.82	41.78	58.22	54.92	0.54	0.67	0.11
MG279121			22.15	23.03	24.56	30.26	9.87	25.66	31.58	32.89	46.71	53.29	56.58	44.74	50.66	41.45	58.55	55.12	0.55	0.68	0.11
MG279122			22.26	22.92	24.45	30.37	10.20	25.66	31.25	32.89	46.71	53.29	56.58	44.74	50.66	41.45	58.55	55.95	0.55	0.68	0.11
MG279123			22.26	22.92	24.45	30.37	10.20	25.66	31.25	32.89	46.71	53.29	56.58	44.74	50.66	41.45	58.55	55.95	0.55	0.68	0.11
MG279125			22.26	22.92	24.45	30.37	10.20	25.66	31.25	32.89	46.71	53.29	56.58	44.74	50.66	41.45	58.55	55.95	0.55	0.68	0.11
MG279126			22.26	23.03	24.56	30.15	10.20	25.66	31.58	32.57	46.82	53.18	56.58	44.74	50.66	41.78	58.22	55.06	0.54	0.68	0.11
MG279127			22.26	22.92	24.56	30.26	10.20	25.66	31.25	32.89	46.82	53.18	56.25	44.74	50.49	41.45	58.55	55.82	0.54	0.68	
MG279128			22.37	23.14	24.23	30.26	10.53	25.66	30.92	32.89	46.60	53.40	56.58	45.07	50.82	41.45	58.55	55.45	0.54	0.69	
MG279129			22.37	23.14	24.23	30.26	10.20	25.99	30.92	32.89	46.60	53.40	56.58	44.74	50.66	41.12	58.88	56.01	0.55	0.69	
MG279130			22.15	22.92	24.56	30.37	10.20	25.66	31.58	32.57	46.71	53.29	56.58	45.07	50.82	41.78	58.22	55.36	0.55	0.68	
MG279131			22.37	22.70	24.67	30.26	10.20	25.33	31.58	32.89	47.04	52.96	56.25	44.41	50.33	41.78	58.22	55.87	0.54	0.68	
MG279135			22.48	22.92	24.34	30.26	10.86	25.33	31.58	32.24	46.82	53.18	56.58	45.39	50.99	42.43	57.57	56.65	0.54	0.69	
MG279140			22.26	22.81	24.56	30.37	10.20	25.33	31.58	32.89	46.82	53.18	56.58	44.74	50.66	41.78	58.22	55.84	0.55	0.68	
MG279141			22.48	22.81	24.45	30.26	10.86	25.33	31.25	32.57	46.93	53.07	56.58	44.74	50.66	42.11	57.89	56.54	0.55	0.68	
MK731981			23.25	23.03	23.68	30.04	13.16	25.66	30.59	30.59	46.93	53.07	58.55	44.41	51.48	43.75	56.25	57.04	0.53	0.71	
MK731982			24.23	22.70	23.36	29.71	14.80	25.33	28.29	31.58	47.59	52.41	56.91	43.42	50.16	43.09	56.91	57.46	0.55	0.72	
MN128702			22.92	23.46	23.46	30.15	12.17	27.63	28.62	31.58	46.38	53.62	57.89	43.75	50.82	40.79	59.21	57.16	0.55	0.71	
MK944079			23.14	23.36	23.25	30.26	12.83	25.66	28.95	32.57	46.38	53.62	58.55	44.08	51.32	41.78	58.22	57.03	0.55	0.72	
MK944080			22.81	23.03	23.79	30.37	12.17	25.33	30.92	31.58	46.60	53.40	59.21	44.08	51.64	43.09	56.91	56.74	0.54	0.70	
KP260925			23.25	22.92	24.01	29.82	12.50	24.67	30.92	31.91	47.26	52.74	56.91	44.74	50.82	43.42	56.58	58.27	0.54	0.70	
KP260926			23.25	23.14	24.01	29.61	12.50	25.00	31.25	31.25	47.26	52.74	56.91	45.07	50.99	43.75	56.25	57.82	0.54	0.71	
KP260927			23.25	23.14	23.90	29.71	12.83	25.00	30.59	31.58	47.15	52.85	56.91	45.07	50.99	43.42	56.58	58.41	0.54	0.71	

KF887949			22.92	23.90	23.03	30.15	12.83	27.63	27.96	31.58	45.94	54.06	57.89	45.07	51.48	40.79	59.21	56.62	0.57	0.73	
KT283604			23.25	24.12	22.81	29.82	12.83	27.96	26.97	32.24	46.05	53.95	56.58	45.07	50.82	39.80	60.20	54.93	0.56	0.74	
KP941114			23.46	22.81	23.36	30.37	47.70	55.92	52.30	42.43	46.82	53.18	74.00	100.00	87.00	98.00	40.00	55.88	0.56	0.71	
MK424788																					
	Average Value ±SD		**27.544**	**24.181**	**22.542**	**25.733**	**24.2**	**23.9**	25.905	26.554	48.968	49.914	53.04	46.894	49.97	50.096	50.121	54.16	0.527	0.83958	0.10977923