Engineering Rotavirus Viral Protein 6 Consensus Sequence to Reduce Aggregation for Expression and Purification

by

Tovhowani Mudau

Submitted in accordance with the requirements for the degree of

MASTERS OF LIFE SCIENCE

in the

COLLEGE OF AGRICULTURE AND ENVIRONMENTAL SCIENCES

DEPARTMENT OF LIFE AND CONSUMER SCIENCE

at the

UNIVERSITY OF SOUTH AFRICA

FLORIDA CAMPUS

SUPERVISOR: Prof. S. Gildenhuys

Co-SUPERVISOR: Dr. N. Parbhoo

September 2025

UNISA | college of agriculture and environmental sciences

Declaration

I, **Tovhowani Mudau**, hereby declare that the dissertation, with the title: "**Engineering Rotavirus Viral Protein 6 Consensus Sequence to Reduce Aggregation for Expression and Purification**" which I hereby submit for the degree of Master of Science in Life Sciences at the University of South Africa, is my own work and has not previously been submitted by me for a degree at this or any other institution.

I declare that the dissertation /thesis does not contain any written work presented by other persons whether written, pictures, graphs or data or any other information without acknowledging the source.

I declare that where words from a written source have been used the words have been paraphrased and referenced and where exact words from a source have been used the words have been placed inside quotation marks and referenced.

I declare that I have not copied and pasted any information from the Internet, without specifically acknowledging the source and have inserted appropriate references to these sources in the reference section of the dissertation or thesis.

I declare that during my study I adhered to the Research Ethics Policy of the University of South Africa, received ethics approval for the duration of my study prior to the commencement of data gathering, and have not acted outside the approval conditions.

I declare that the content of my dissertation/thesis has been submitted through an electronic plagiarism detection program before the final submission for examination.

Student signature:  **Date:** 30 September 2025

Abstract

The VP6 protein sequence from Rotavirus A was analysed using AGGRESCAN3D to identifying aggregation-prone regions. Six amino acid residues (four Valine, one Phenylalanine, and one Isoleucine) were mutated to reduce aggregation. VP6 consensus (wild-type) and VP6 mutant genes were cloned into pET-15b for expression with BL21(DE3) and NiCo21(DE3) using different conditions. NiCo21(DE3) yielded the highest expression in inclusion bodies at 37°C, 0.04 mM IPTG, after 12 hours. An optimized wash step was added to improve purity, and optimal solubilisation was achieved with 7 M urea for VP6 wild-type protein, while the VP6 mutant protein required 5 M and 7 M urea, demonstrating significantly improved solubility, following a freeze-thaw cycle. Purification by Immobilized Metal Affinity Chromatography (IMAC) showed VP6 co-eluting with *E. coli* proteins confirmed by mass spectrometry: EF-Tu and PheS with VP6 wild-type protein, and YcfT and MalE with VP6 mutant protein. Structural analysis using Far UV Circular Dichroism (CD), Attenuated Total Reflectance–Fourier Transform Infrared Spectroscopy (ATR-FTIR), and fluorescence spectroscopy, showed a complete loss of secondary and tertiary structure at 7 M urea, whereas the VP6 mutant protein showed significant unfolding at 5 M urea, indicating reduced structural stability. Thermal denaturation monitored by Far-UV circular dichroism and intrinsic fluorescence showed consistent unfolding behaviour for VP6 proteins. The VP6 wild-type protein began unfolding at 35°C, while the VP6 mutant protein showed a greater thermal stability, initiating unfolding at 45°C. Despite this difference, neither protein demonstrated refolding upon cooling, indicating irreversible denaturation. Native-PAGE showed VP6 forming higher-order quaternary structures. The VP6 wild-type protein formed multiple oligomeric bands, while the VP6 mutant protein formed fewer, weaker bands, indicating reduced oligomeric complexity and disrupted quaternary stability.

Key words: Rotavirus A, VP6, Protein engineering, *Escherichia coli* expression, inclusion bodies, Solubilisation, immobilised metal affinity chromatography

Research Outputs

1.Poster Presentation at UNISA 2024

Engineering a More Soluble Viral Protein 6 Consensus Sequence for Improved Expression and Purification.

Mudau, T, Parbhoo, N, and Gildenhuys, S.

2. Manuscript in Progress

Structure-Guided Mutagenesis and Biophysical Analysis of Rotavirus A VP6 Reveal Effects on Protein Aggregation

Mudau, T, Parbhoo, N, and Gildenhuys, S.

Dedication

I dedicate this work to my late father, Mr. Johannes Pharanndwa Mudau. I wish you were here to witness this moment me completing my MSc, the first in our family to achieve this milestone. To my mother, Mrs. Elsie Ntsieni Mudau, thank you for your constant encouragement and patience. It took time, but I finally made it, just as you always hoped I would. To my sister, Ms. Awelani Mudau, u tambula hanga no huvhona ndisa edeli, ndikho livhuwa thuthuwedzo, thikhedzo nau dzula ni hone musi nditshi toda thuthuwedzo na ngeletshedzo.

Acknowledgement

I am immensely thankful to:

Family: I am deeply grateful to my family for their unwavering support, encouragement, and love throughout this journey. Their belief in me has been a constant source of strength and motivation. Without their patience and understanding, completing this work would not have been possible.

Ms Tracy-Lee Vermaak: I am incredibly thankful; you were my lab partner and a sister in this journey. We endured the challenges of our master's studies side by side, supporting each other through every tough moment. Your friendship, encouragement, and shared struggles made this experience truly unforgettable.

Ms. Thato Manyaapelo: I am sincerely grateful to you for your valuable support throughout my research. Your positive attitude and expert assistance with the equipment, patience, and willingness to help ensured smooth progress in the lab.

Mr Tshepo Sekele: My heartfelt thanks for your support and encouragement. Your expertise and collaborative spirit provided valuable input and motivation throughout this process.

Mr Garland Moore: Your generosity in granting me access to the Nicolet iS5 ATR-FTIR instrument was invaluable to the completion of my research. I am deeply grateful for your support and assistance.

Dr. Nishal Parbhoo- My co-supervisor: I am truly thankful, you were my go-to person during moments of stress, confusion, and self-doubt. Your constant support, kindness, and understanding kept me grounded and motivated. I am deeply grateful for your unwavering guidance and encouragement.

Professor Samantha Gildenhuys - My supervisor: Thank you for your exceptional guidance. Your wisdom, support, and belief in my abilities were vital throughout this journey. I deeply appreciate your time and efforts invested in helping me succeed.

The National Research Foundation (NRF): I am truly thankful for the financial support towards my postgraduate journey.

Table of Contents

Declaration i
Abstract ii
Research Outputs iii
Dedication iv
Acknowledgement v
List of Figures ix
List of Table xi
Summary xii
List of Abbreviations xiv
List of Buffers xvi
1 Introduction and Literature Review 1
1.1 Rotavirus Burden and Epidemiology 1
1.2 Rotavirus Vaccine 1
1.3 Rotavirus Structure 3
1.4 Rotavirus Replication Cycle 5
1.5 Rotavirus Classification 7
1.6 Viral protein 6 (VP6) 7
1.7 Expression and Solubilisation of the Rotavirus VP6 Protein 10
1.8 Purification of Rotavirus VP6 Protein 12
1.9 Computational Approach for Protein Modelling and Optimisation 14
1.10 VP6 Protein Structural Conformation and Stability 16
2 Research Aim and Objectives 21
3 Methodology 22
3.1 Materials 22
3.2 Protein Consensus Sequence Identification and Engineering 22
3.3 Bacterial cells Transformation and Cells selection 25
3.3.1 *Escherichia coli* JM109 Cells Competency 25
3.3.2 *Escherichia coli* BL21(DE3) and NiCo21(DE3) Competency 25
3.3.3 Bacterial Cells Transformation 26
3.4 Glycerol Stocks 26
3.5 DNA Extraction 27
3.6 Bacterial Growth Curves and VP6 Protein Expression 28
3.6.1 Bacterial Growth Curve 28

3.6.2 VP6 Protein Expression 28
3.7 Isolation of the VP6 Protein Inclusion Bodies 29
3.8 VP6 Protein Solubilisation 30
3.9 VP6 Protein Purification 31
3.9.1 Affinity Chromatography 31
3.9.2 Analysis of Protein Purity Post-Purification 35
3.10 Gel electrophoresis of the VP6 proteins 36
3.10.1 SDS-PAGE Electrophoresis 36
3.10.2 VP6 Quaternary Structure 37
3.10.2.1 Computational Prediction of the VP6 protein 37
3.10.2.2 Native-PAGE 37
3.11 Purified VP6 Proteins Buffer Exchange and Concentration Determination 38
3.11.1 Buffer Exchange (Dialysis) 38
3.11.2 Buffer Exchange (Desalting) 40
3.11.3 Concentration Determination 40
3.12 Spectroscopy 42
3.12.1 Mass Spectrometry 42
3.12.1.1 In-gel Digestion 42
3.12.1.2 LC-MS Analysis 42
3.12.1.3 Data Analysis 43
3.12.2 Characterization of the Secondary Structure 43
3.12.2.1 Circular dichroism 43
3.12.2.2 Attenuated Total Reflection (ATR)–Fourier Transform Infrared (FTIR) Spectroscopy 44
3.12.3 Intrinsic Tryptophan Fluorescence 45
3.12.4 Thermal Unfolding 45
4 Results 47
4.1 Protein Consensus Sequence Identification and Engineering 47
4.2 Bacterial Growth Curve and VP6 Protein Overexpression 51
4.2.1 Bacterial Growth Curve 51
4.2.2 VP6 Protein Expression 55
4.2.2.1 Overexpression of the VP6 Wild-type Protein 58
4.2.2.2 Overexpression of the VP6 Mutant Protein 61
4.3 Isolation of the VP6 Protein Inclusion Bodies and Solubilisation 64
4.4 The VP6 Protein Solubilisation 66

4.5 VP6 Protein Affinity Chromatography Purification 68
4.5.1 VP6 Wild-type Protein Purification 68
4.5.2 VP6 Mutant Protein Purification 77
4.6 VP6 Protein Concentration Determination 82
4.7 Mass Spectroscopy Analysis 85
4.8 Characterisation of the Secondary Structure 89
4.8.1 VP6 Wild-type Protein Secondary Structure 89
4.8.2 VP6 Mutant Protein Secondary Structure 93
4.9 Intrinsic Fluorescence 98
4.10 Thermal Unfolding of the VP6 Protein 102
4.10.1 Far-UV Circular Dichroism Thermal Unfolding 103
4.10.2 Fluorescence Monitoring of Rotavirus VP6 Thermal Unfolding 109
4.11 Quaternary Structure Characterization 115
5 Discussion 119
6 Conclusion 123
7 Future Directions 124
8 References 125

List of Figures

Figure 1.1: Rotavirus Structure 4
Figure 1.2: Rotavirus Replication Cycle 6
Figure 1.3: VP6 Protein Structure 8
Figure 1.4: VP6 Protein Tryptophan Positions 20
Figure 4.1: VP6 Consensus Sequence Optimisation 48
Figure 4.2: VP6 Mutated Sequence and Aggrescan3D Score 49
Figure 4.3: VP6 Protein Structural Prediction 50
Figure 4.4: BL21(DE3) with VP6 Encoding Plasmid Growth Curve 53
Figure 4.5: NiCo21 (DE3) with VP6 Encoding Plasmid Growth Curve 54
Figure 4.6: VP6 Wild-type Protein Molecular Weight Calculation 56
Figure 4.7: VP6 Mutant Protein Molecular Weight Estimation 57
Figure 4.8: VP6 Wild-type Protein Expression using BL21 (DE3) 59
Figure 4.9: VP6 Wild-type Protein Expression using NiCo21 (DE3) 60
Figure 4.10: VP6 Mutant Protein Expression using BL21 (DE3) 62
Figure 4.11: VP6 Mutant Protein Expression Using NiCo21 (DE3) 63
Figure 4.12: Isolation of Inclusion Bodies for VP6 Protein 65
Figure 4.13: VP6 Wild-type and VP6 Mutant Protein Solubilisation 67
Figure 4.14: VP6 Wild-type Protein Purification Trial 1 70
Figure 4.15: VP6 Wild-type Protein Purification Trial 2 71
Figure 4.16: VP6 Wild-type Protein Purification Trial 3 73
Figure 4.17: VP6 Wild-type Protein Purification Trial 4 74
Figure 4.18: VP6 Wild-type Protein Purification Trial 5 76
Figure 4.19: VP6 Mutant Protein Purification Trial 1 78
Figure 4.20: VP6 Mutant Protein Purification Trial 2 80
Figure 4.21: VP6 Mutant Protein Purification Trial 3 81
Figure 4.22: Purified VP6 Protein Absorbance Spectrum 84
Figure 4.23: Mass Spectrometry Analysis of VP6 Wild-type Protein 86
Figure 4.24: Mass Spectrometry Analysis of VP6 Mutant Protein 88
Figure 4.25: VP6 Wild-type Protein Secondary Structure Characterisation 90
Figure 4.26: VP6 Mutant Protein Secondary Structure Characterisation 94
Figure 4.27: Intrinsic Fluorescence Spectra of the VP6 Wild-type Protein 99
Figure 4.28: Intrinsic Fluorescence Spectra of the VP6 Mutant Protein 101
Figure 4.29: VP6 Wild-type Protein Far UV Thermal Unfolding 104
Figure 4.30: VP6 Mutant Protein Far UV Thermal Unfolding 107
Figure 4.31: VP6 Wild-type Protein Thermal Unfolding 110

Figure 4.32: VP6 Mutant Protein Thermal Unfolding 113
Figure 4.33: VP6 Protein Quaternary structure 116

List of Table

Table 3.1: VP6 Protein Purification Overview..... 32

Summary

Rotavirus (RV) is a pathogenic virus characterized by its double-stranded segmented RNA genome and is a member of the family *Reoviridae*. Rotavirus infection remains a significant cause of gastroenteritis in children less than 5 years old throughout the world. Rotaviruses are non-enveloped viruses composed of triple-layered viral particles with VP6 as the middle layer. VP6 protein shows promise as a vaccine candidate, a nano-delivery mechanism, and an adjuvant for fused proteins. This study aimed to optimise the VP6 sequence to improve the *Escherichia coli (E. coli)* expression, solubility, and purification of the VP6 protein. The VP6 protein sequences of Rotavirus A from 2017–2024 were retrieved from the National Center for Biotechnology Information (NCBI) database and aligned using Jalview to generate a consensus sequence. AlphaFold (via ColabFold) was used to predict the 3D structure of the consensus sequence, producing five models; the optimal structure was selected for analysis. Aggregation-prone regions were identified using the AGGRESCAN3D and they were optimized using amino acid substitutions. Six residues (Valine 39 mutated to Alanine, Valine 109 mutated to Leucine, Valine 151 mutated to Threonine, Phenylalanine 248 mutated to Histidine, Isoleucine 252 mutated to Threonine, and Valine 281 mutated to Threonine) were mutated based on their aggregation scores to reduce aggregation while maintaining potential naturally occurring substitutions. The Expasy ProtParam tool was used to analyse the VP6 consensus (wild-type) sequence and VP6 mutant sequence, then the sequences were reverse translated and the gene encoding the VP6 proteins was codon optimised and cloned into pET-15b for expression. The *E. coli* strains BL21(DE3) and NiCo21(DE3) were used for the expression study. The VP6 proteins were expressed using different expression conditions, including temperatures (37°C and 20°C), isopropyl-β-D-thiogalactopyranoside (IPTG): (BL21 (DE3) IPTG concentration was 1 mM and 0.1 mM: NiCo21 (DE3) IPTG concentration was 0.4 mM and 0.04 mM) and the post induction time was from 0 hours and 4 hours to 12 hours. SDS-PAGE analysis showed VP6 proteins overexpression in inclusion bodies in both *E. coli* strains with a small amount of soluble protein detected in the supernatant. However, NiCo21(DE3) demonstrated the highest yield of both VP6 wild-type protein and VP6 mutant protein after 12 hours post-induction at a temperature of 37°C using 0.04 mM IPTG concentration. An optimized wash step was implemented during to enhance the removal of non-specifically bound host proteins to improve the protein purity.

To fully recover the functional VP6 proteins in inclusion bodies an optimal solubilisation was successfully achieved at 5 M, and 7 M urea concentrations at different pH (pH 7.5 and 8) after a freeze-thaw cycle. Different trials of the immobilised metal affinity chromatography (IMAC) were conducted to purify both the VP6 wild-type protein and VP6 mutant protein with different buffer composition which included different concentrations of urea, NaCl, and imidazole, as well as the Tween-20. Despite the optimisation steps, SDS-PAGE analysis revealed the presence of two faint bands co-eluting with the VP6 proteins, likely due to non-specific interactions with the His-tag. Mass spectrometry identified these bands as *E. coli* proteins. For the VP6 wild-type protein, the co-eluting *E. coli* proteins were EF-Tu and PheS, while for the VP6 mutant protein, the co-eluting contaminants were identified as YcfT and MalE. After the VP6 protein confirmation, their structural features or characteristics were quantified. The Far-UV circular dichroism and ATR-FTIR spectroscopy were used to confirm the presence of secondary structure in different urea concentrations (0 M, 4 M, 5 M and 7 M) showing a partially folded structure with both alpha-helices and beta-sheets at 4 M urea and a red shift and intensity changes in specific peaks, indicating alterations in secondary structure as the urea increases to 7 M. The Intrinsic fluorescence spectroscopy was excited at 295 nm and when the urea concentration increased from 4 M there was a red shift consistent with tertiary structure disruption and solvent exposure of aromatic residues of the VP6 proteins. The thermal unfolding (far UV Circular Dichroism and Intrinsic fluorescence spectroscopy) showed irreversible denaturation on both the VP6 protein from 20°C to 94°C, cooling back to 20°C and overnight at 4°C, with VP6 wild-type protein melting temperature at 35°C and VP6 mutant protein melting point at 45°C. Denaturation was marked by a progressive loss of secondary and tertiary structures with minimal recovery upon cooling, indicative of irreversible aggregation or misfolding. Native-PAGE was used to assess quaternary assembly, which showed a higher order of different structures or species.

List of Abbreviations

ATR-FTIR: Attenuated Total Reflection–Fourier Transform Infrared Spectroscopy

BSA: Bovine Serum Albumin

DLPs: double-layered particles

dsRNA: Double-stranded RNA

E. coli: *Escherichia coli*

ER: Endoplasmic Reticulum

Far-UV CD: far-ultraviolet circular dichroism

F: Fraction

IMAC: immobilised metal affinity chromatography

pI: isoelectric point

IPTG: Isopropyl β-d-1-thiogalactopyranoside

LB: Luria-Bertani

mAbs: Monoclonal antibodies

MBP: Maltose Binding Protein

MCS: Multiple cloning sites

MMseqs2: Many-against-Many sequence searching

MWM: Molecular weight marker

NCBI: National Center for Biotechnology Information

NSP: non-structural RNA-binding protein

OD: Optical Density

PDB: Protein Data Bank

RV: Rotavirus

RVA: Rotavirus A

SF: Sample flow-through

SG: Subgroup

SOB: Super Optimal Broth

TLPs: Triple Layered Particles

UV: Ultraviolet

VLPs: Virus-Like Particles

VP: Viral Proteins

WHO: World Health Organization

International Union of Pure and Applied Chemistry-International Union of Biochemistry and Molecular Biology (IUPAC-IUBMB) one-letter and three-letter codes for amino acids were used.

List of Buffers

Lysis buffer: 50 mM Tris-HCl, pH 7.5 and pH 8.0, 150 mM NaCl, and 0.02% (w/v) sodium azide.

Solubilisation Buffer A: 50 mM Tris-HCl, pH 7.5 and pH 8.0, 500 mM NaCl, and 0.02% (w/v) sodium azide.

Solubilisation Buffer B: 50 mM Tris-HCl, pH 7.5 and pH 8.0, 150 mM NaCl, 2 M, 5 M, 7 M Urea, and 0.02% (w/v) sodium azide.

Wash Buffer A: 50 mM Tris-HCl. pH 7.8, and 1% TritonX-100.

Wash Buffer B: 50 mM Tris-HCl, pH 7.8.

Wash Buffer C: 50 mM Tris-HCl pH 7.5 and pH 8.0, 5 M urea, 150-300 mM NaCl, 150 mM Imidazole, and 0.02% (w/v) sodium azide

Wash Buffer D: 50 mM Tris-HCl pH 7.5, 7 M urea, 150 mM NaCl, 80 mM Imidazole, 0.5% Tween-20, and 0.02% (w/v) sodium azide

Wash Buffer E: 50 mM Tris-HCl pH 7.5, 7 M urea, 300 mM NaCl, 80 mM Imidazole, 0.5% Tween-20, and 0.02% (w/v) sodium azide

Equilibration Buffer A: 50 mM Tris-HCl pH 7.8, 5 M urea, 300 mM NaCl, 0-20 mM Imidazole, and 0.02% (w/v) sodium azide.

Equilibration Buffer B: 50 mM Tris-HCl pH 7.8, 5 M urea, 150 mM NaCl, 50 mM Imidazole, and 0.02% (w/v) sodium azide.

Equilibration Buffer C: 50 mM Tris-HCl pH 7.5, 7 M urea, 150mM NaCl, 80mM Imidazole, and 0.02% (w/v) sodium azide.

Elution Buffer A: 50 mM Tris-HCl pH 7.5 and pH 8.0, 5M urea, 300 mM NaCl, 500-600 mM Imidazole, and 0.02% (w/v) sodium azide.

Elution Buffer B: 50 mM Tris-HCl pH 7.5, 7 M urea, 150 mM NaCl, 600 mM Imidazole, and 0.02% (w/v) sodium azide.

Dialysis buffer A: 50 mM sodium phosphate dibasic, pH 7.5, 7 M, 4 M, 1 M, and 0 M urea, and 0.02 % (w/v) sodium azide.

Dialysis buffer B: 50 mM sodium phosphate dibasic, pH 7.5, 20 mM NaCl, 7 M, 4 M, 1M, 0 M urea, and 0.02 % (w/v) sodium azide.

Dialysis buffer C: 50 mM sodium phosphate dibasic, pH 5.5 and pH 8, 20 mM NaCl, 7 M, 4 M, 1M, and 0 M urea, and 0.02 % (w/v) sodium azide.

Dialysis buffer D: 50 mM sodium phosphate dibasic, pH 7.5 containing 20 mM sodium chloride, 10% glycerol, 7 M, 4 M, 2 M, and 0 M urea, and 0.02 % (w/v) sodium azide.

Dialysis buffer E: 50 mM sodium phosphate dibasic, pH 7.8 containing 20 mM sodium chloride, 7 M, 4 M, 4 M urea, and 0.02 % (w/v) sodium azide.

Dialysis buffer F: 50 mM sodium phosphate dibasic, pH 7.8 containing 20 mM sodium chloride, 0 M urea, and 0.02 % (w/v) sodium azide.

1 Introduction and Literature Review

1.1 Rotavirus Burden and Epidemiology

Rotavirus (RV) is characterized by its double-stranded segmented RNA genome, recognized as a leading cause of gastroenteritis in children under the age of five years old throughout the world, and it is a member of the family *Reoviridae* (Abid *et al.* 2019). Almost every child in the world experiences an infection with rotavirus by the age of five years old, with most developing gastroenteritis as a result (WHO, 2021). The most known RV infection complication that occurs at a young age is the risk of dehydration, which can be linked to metabolic acidosis and electrolyte deficiency. Dehydration can be managed with oral or intravenous fluid replacement (Dennehy, 2015; Smok *et al.* 2016). Diarrhoea is estimated as the second leading cause of death among young children and infants, with over 1.7 billion cases of rotavirus, with nearly 525 000 deaths occurring globally every year (Hartman *et al.* 2022). In South Africa, prior to the introduction of the rotavirus vaccine, rotavirus infection was reported to cause an estimated 17,644 to 25,630 hospitalisations (Asowata *et al*. 2018; Mapaseka *et al*. 2010). Rotavirus infection remains a significant burden in developing countries, where infection, hospitalization, and death rates among children under five years old remain notably higher compared to developed regions such as the United States and Europe (Karampatsas *et al.* 2018).

1.2 Rotavirus Vaccine

Prevention is a key strategy in controlling rotavirus infections. Although the first rotavirus vaccine became available in the United States in 1999, it was withdrawn just over a year later due to concerns about it, linked to intussusception (Dennehy, 2008). Currently, the World Health Organization (WHO) has approved four rotavirus vaccines: including Rotarix®, RotaTeq®, Rotasiil®, and Rotavac® (Burke *et al.* 2019; WHO, 2021). Out of the four vaccines, Rotasiil and Rotavac are used in India, while RotaTeq and Rotarix are used globally (Zade *et al.* 2014). RotaTeq, a bovine-human reassortant pentavalent vaccine, includes genotypes G1P[5], G2P[5], G3P[5], G4P[5], and G6P[8], which are considered globally significant due to their prevalence among circulating rotavirus strains (Burke *et al.* 2019; Vesikari *et al.* 2004).

Rotarix is a monovalent vaccine consisting of a single human rotavirus strain, G1P[8], which accounts for approximately 39% of circulating RVA infections in Africa (Omatola *et al.* 2024)

RotaTeq and Rotarix were first introduced in Europe and America in 2006. The vaccine's initial implementation in Africa began with South Africa in 2009, paving the way for its rollout across other regions, especially in similar socio-economic contexts (Seheri *et al.* 2012; Hallowell *et al.* 2020). As of the last review in November 2019, 42 out of 54 African countries had successfully implemented rotavirus immunization programs (ROTA Council, 2022). Countries that implemented the immunization program included the 27 countries that financially depend on Gavi, which is part of the vaccine union (Mwenda *et al.* 2018). Countries like South Africa, Ghana, Malawi, and Rwanda introduced the rotavirus vaccines (RotaTeq and Rotarix) earlier and have observed their effectiveness, with a reduction in hospitalizations for rotavirus-related diseases by 40%–60% in young children and infants, the decline in hospitalizations was even more substantial, with reductions ranging from 50% to 70%. (Mwenda *et al.* 2018; Steele and Groome, 2019; Bar-Zeev *et al.* 2016). Reported data from the United States, Latin America, and Europe indicated that the vaccine prevents rotavirus illness with an effectiveness rate ranging from 85% to 98% (Kirkwood *et al.* 2019).

In developing countries, the rotavirus vaccines do not exhibit the same level of effectiveness as they do in developed countries, highlighting the urgent need for improved vaccine production methods or innovative new strategies to enhance their impact (Carcamo-Calvo *et al.* 2021). Vaccine development against rotavirus has focused on structural proteins VP4, VP7, and VP6, with VP4 and VP7 recognized for inducing neutralizing antibody responses (Dennehy, 2008). There has also been growing interest in VP6 as a potential vaccine candidate due to its ability to generate immunoglobulin antibodies, making it a promising option for vaccines targeting various viral strains (Afchangi *et al.* 2019). Because their safe, immunogenic, and non-infectious properties, Virus-like particles (VLPs) are also being studied as promising vaccine candidate. (Yong *et al*. 2004). According to Jere *et al.* (2014), virus-like proteins can be used as candidates for non-live vaccine production against rotavirus strains in humans and animals.

1.3 Rotavirus Structure

A defining feature of rotaviruses is their triple-layered particle structure (TLP) and lack of an outer envelope (Ward and McNeal, 2010). These particles consist of an outer, intermediate, and inner layer (Figure 1.1). The virus's outer layer is composed of VP4, which forms spikes from 60 trimers, and VP7, composed of 260 trimers. Both proteins play essential roles in cell entry and help determine rotavirus serotypes. Additionally, they promote the humoral immune response by inducing neutralizing antibodies (Ludert *et al.* 2002). The cleavage of VP4 by trypsin into VP5 and VP8 (Figure 1.1), enhances viral infectivity (Ludert *et al.* 2002; Dormitzer *et al.* 2001). This cleavage facilitates supports for viral penetration but does not affect binding to the cells. VP6, the major capsid protein, forms the intermediate layer of the rotavirus particle, consisting of 260 trimers. VP6 acts as a subgroup antigen, used to classify rotavirus A strains by groups and subgroups (Mathieu *et al.* 2001). During the rotavirus replication cycle, as rotavirus enters the host cell, the VP4 and VP7 (outer proteins) are removed, leaving a double-layered particle (DLP) with VP6 forming the new outer layer. The integrity of this non-infectious double-layer particle is crucial during transcription, highlighting the importance of VP6 in the virus replication process (Estes *et al*. 2001). The inner layer consists of 60 asymmetric dimers of VP2 (120 molecules), which enclose the enzymes VP1 and VP3, as well as 11 segments of double-stranded RNA (dsRNA) (Figure 1.1). Within the DLP, VP1 functions as the polymerase that produces single-stranded positive-sense RNAs (+RNAs), which are subsequently capped at the 5' end through the enzymatic activity of VP3 (Lawton *et al.* 2000; Patton *et al*.1997). VP1 uses the +RNAs as templates for translating viral proteins and replicating the genome. Out of the 11 dsRNA genome segments, 10 encode for single proteins, while the eleventh segment encodes two non-structural proteins, NSP5 and NSP6 (Estes *et al.* 2001). Out of the 12 genome-encoded proteins, six are structural (VP1, VP2, VP3, VP4, VP6, and VP7) and six are non-structural (NSP1, NSP2, NSP3, NSP4, NSP5, and NSP6) (Greenberg and Estes, 2009; Desselberger, 2014; Estes *et al*. 2007; Long and McDonald, 2017). Each plays a critical role during the rotavirus replication cycle.

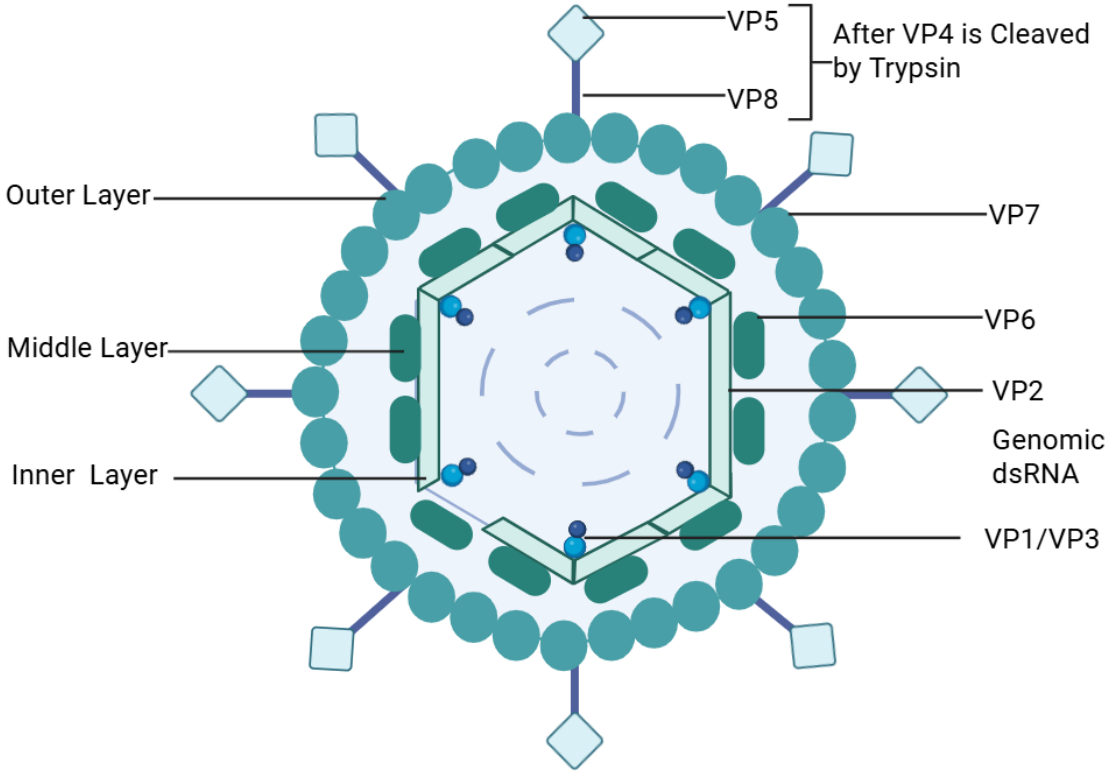


Figure 1.1: Rotavirus Structure

The figure was constructed from https://biorender.com, it shows triple-layered particles represented in different shapes (rectangle for the inner layer, oval for the middle layer, and circle for the outer layer). VP4 and VP7, with VP4 being cleaved by trypsin into VP5 and VP8. The middle layer is made up of VP6. The inner layer is made up of VP2 which surrounds the enzymes VP1 and VP3 as well as the 11 segments of double-stranded RNA (dsRNA) genome.

1.4 Rotavirus Replication Cycle

During cell entry, the TLPs bind to host cell receptors (Lawton *et al.* 2000; Long and McDonald, 2017; Arias and Lopez, 2021), and the outer layers (VP4 and VP7) are lost during endocytosis. As a result, transcriptionally active DLPs are formed (Figure 1.2) (Settembre *et al.* 2011). The positive-strand RNA molecules, which can function as mRNA, are transcribed inside the DLP, allowing the dsRNA strand to exit through the aqueous channel from the VP2 to the VP6 layers of the DLPs (Charpilienne *et al.* 2002). The integrity of VP6 is essential for mRNA transcript production. These mRNAs are translated into 12 viral proteins: six non-structural proteins (NSP1, NSP2, NSP3, NSP4, NSP5, and NSP6) and six structural proteins (VP1, VP2, VP3, VP4, VP6, and VP7). following translation, the viral proteins enable the replication of RNA genomes, which are then incorporated into new DLPs formed inside the viroplasm. Notably, NSP2 and NSP5, play a crucial role in the development of the viroplasm (Papa *et al*. 2021). Inside the viroplasm, NSP2, NSP5, and NSP6 adapt and alter the cellular machinery to manage various stages of replication, including protein assembly, RNA replication, and DLP formation, after the initial transcripts are translated (Ruiz *et al*. 2009; Hu *et al.* 2012).

The newly formed DLPs, containing VP6, VP3, VP2, VP1, and the 11 RNA segments, are released from the viroplasm and bind to the NSP4, which functions as a receptor to initiate the assembly of TLPs (Figure 1.2) (Crawford *et al*. 2019). The NSP4 also functions as a viroporin, which facilitates the calcium release from the internal cell stores, which is required for outer layer assembly (Desselberger, 2014). The DLPs containing the translated proteins VP2 and VP6 enter the endoplasmic reticulum (ER), and a temporary envelope is formed. The newly synthesized translated outer layer proteins VP4 and VP7 assemble into the DLPs, displacing the temporary envelope and resulting in the formation of mature, infectious TLPs. These mature virions exit the host cell through lysis and escape via exocytosis (Crawford *et al.* 2019).

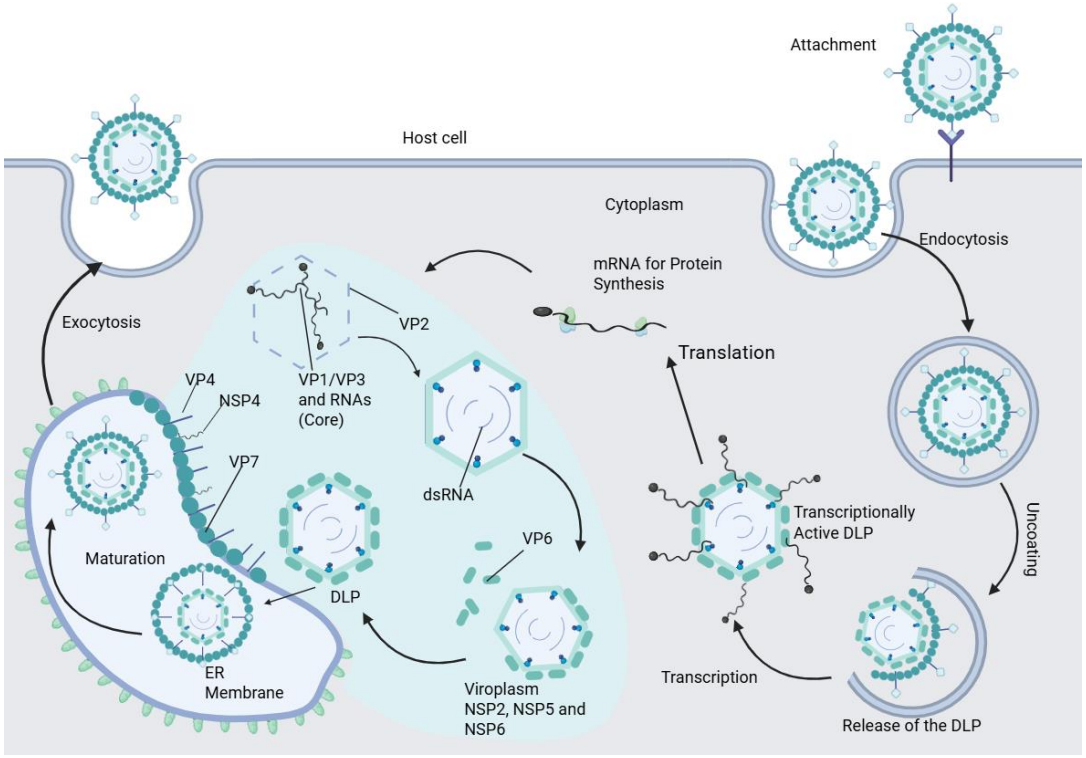


Figure 1.2: Rotavirus Replication Cycle

Adapted from (Crawford *et al.* 2017: Desselberger, 2014), the figure was created using Biorender Illustration Software https://biorender.com. The figure shows how the replication cycle of the rotavirus occurs, from cell entry, formation of double-layered particles to the translation of protein into forming a new triple-layered particle and releasing the viral particles via cell lysis.

1.5 Rotavirus Classification

Rotaviruses are categorized into groups, subgroups, and serotypes, primarily based on the RVA, VP4 and VP7 outer capsid proteins, which exhibit independent antigenic specificities (Hoshino *et al.* 1985). The VP7, a glycoprotein, determines the specificity of the G serotype (because VP7 is Glycoprotein 36G), which corresponds with the G genotype, while VP4, a protease-sensitive protein, defines the P serotype (because VP4 is protease sensitive 51P). Notably, more G genotypes than P serotypes have been identified (Taniguchi and Urasawa, 1995; Donato and Bines, 2021). Currently, nine rotavirus groups (A–J) have been recognized by the International Committee on Taxonomy of Viruses (ICTV) (Crawford *et al.* 2017; Matthijnssens *et al*. 2012). Groups A, B, C, and H infect both humans and animals, while groups D, F, G, I, and J infect only animals and birds (Giaquinto *et al.* 2007; Mihalov-Kovács *et al.* 2015). Additionally, there are also two recently identified groups K (RVC-like) and L (RVH-like) (Johne *et al.* 2019; Johne *et al.* 2022; Johne *et al.* 2023).

The VP6 protein, located in the intermediate layer, carries antigenic epitopes that enable the classification of group A rotaviruses (RVA) into four subgroups: SG I, SG II, SG I/II, and SG non-I/non-II (Greenberg *et al.* 1983; Van Damme *et al.* 2007). Identification of groups and subgroups is based on VP6's ability to bind to specific monoclonal antibodies (mAbs) and the electrophoretic mobility patterns of the virus's 11 dsRNA genome segments (Estes *et al*. 2001; Greenberg *et al.* 1983).

1.6 Viral protein 6 (VP6)

The VP6 protein forms trimeric structures that make up the intermediate layer of the rotavirus particle (Figure 1.3), situated between the outer layer (VP4 and VP7) and the inner layer (VP2). Two domains make up each VP6 monomer: an alpha-helical domain and a beta-sheet domain (Mathieu *et al.* 2001). The alpha-helical domain is formed by five helices composed of amino acid residues located at the N-terminal (1–150) and the C-terminal (335–397) region (Charpilienne *et al.* 2002; Mathieu *et al.* 2001). The alpha-helical domain is involved in the interaction with the VP2 (inner layer) of the virion, which assists in the build-up of the virus structure ensuring proper viral RNA transcription process during the formation of the rotavirus's active core.

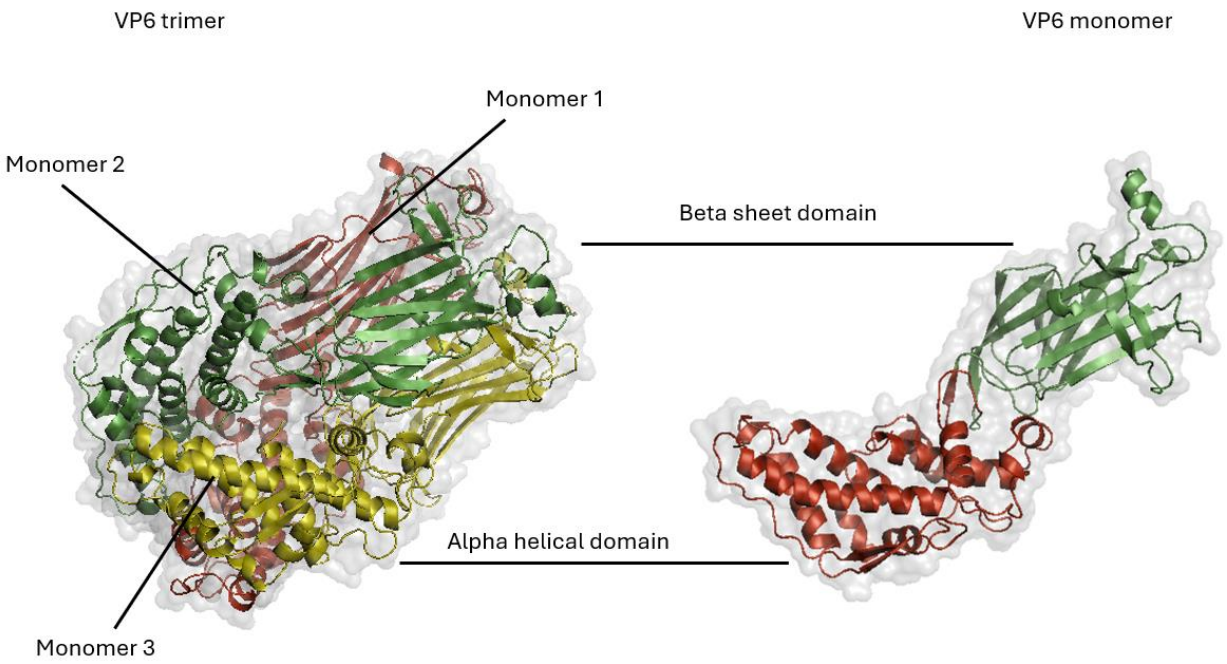


Figure 1.3: VP6 Protein Structure

The biological assembly of the VP6 protein, which forms a trimer composed of three identical monomers highlighted in different colours (red shows monomer 1, green shows monomer 2 and yellow shows monomer 3). The VP6 monomer structure contains the alpha helical domain (red region) and the beta sheet domain (green region). The structural diagrams were generated using PyMOL (The PyMOL Molecular Graphics System, Version 3.0 Schrödinger, LLC). The VP6 PDB structure code: 1QHD (Mathieu *et al.* 2001).

The beta-sheet domain is formed by amino acid residues 151–334 and it is distal from the centre or core of the VP6 trimeric structure (Mathieu *et al.* 2001). The beta-sheet domain interacts with VP7, the outer layer of the virion, and contributes to the virus's ability to infect human cells. These domains work together in shaping and stabilizing the protective shell of the virus and assist in the infection and replication of the virus. A clear understanding of the VP6 protein domains is essential for targeting their functions and disrupting them to potentially develop novel antiviral therapies. Making up about 39% of the total viral particle mass, the VP6 protein is highly immunogenic and capable of self-assembling into various structures, including spherical, tubular, or sheet-like structures, depending on environmental factors such as pH, ionic strength, and the presence of calcium or magnesium ions (Teng *et al.* 2014; Shoja *et al.* 2022). Due to its immunogenicity and self-assembly properties, VP6 is considered a promising antigen and a structural platform for the development of non-living rotavirus vaccines, as well as a nanostructure for biomedical applications (Shoja *et al.* 2022). Additionally, VP6 can function as an immune carrier or adjuvant system and has potential as a delivery vehicle for therapeutic agents.

Several platforms are currently being explored for the development of VP6-based rotavirus vaccines, including the VP6 recombinant protein vaccine and the VP6 VLPs vaccine. Both of which demonstrated protective immunity against rotavirus infection in animal studies (Afchangi *et al.* 2019). Heinimaki *et al.* (2020) identified VP6 nanostructures as a safe and immunogenic vaccine candidate. This was also supported by Parbhoo *et al.* (2016), who reported that the VP6 gene is highly conserved among rotavirus strains, supporting its potential as a broadly protective vaccine target. Ward and McNeal, (2010) further explored the potential of VP6 by investigating its effectiveness in various mouse models. In their study, VP6 was expressed as a chimera with maltose-binding protein (MBP: VP6) and administered intranasally, demonstrating its viability as a mucosal vaccine candidate. When mice were infected with rotavirus, they were nearly completely protected from infection, demonstrating the effectiveness of the VP6 antigen as a vaccine candidate. The VP6 recognition site contains charged quaternary epitopes that are fully exposed and negatively charged.

Aiyegbo *et al*. (2013) showed that human VP6-specific monoclonal antibodies bind to a quaternary, negatively charged epitope located at the transcriptional pore (Type I

channel) of transcriptionally active rotavirus particles. By blocking this channel, these antibodies prevent nascent viral mRNAs from exiting the particle and thereby inhibit transcription and replication. Given its important structural and functional roles in the rotavirus replication cycle, including formation of the middle capsid layer and participation in transcriptional pores, VP6 is considered an essential antigen, and immunity to VP6 has the potential to interfere with viral transcription and replication. Although the native VP6 lacks the major neutralizing epitopes present on the outer capsid proteins (VP4/VP7) and has limited natural insertion sites for foreign epitopes, it is still considered a suitable vector for presenting foreign epitopes (Zhao *et al.* 2011).

1.7 Expression and Solubilisation of the Rotavirus VP6 Protein

Several expression platforms have successfully produced VP6 protein, including *Escherichia coli* (*E. coli*) (Zhao *et al.* 2011), baculovirus-silkworm systems (Yao *et al*. 2012; da Silva Junior *et al.* 2012), *E. coli* and yeast (Bredell *et al.* 2016) and mammalian cells (Choi *et al.* 2002). Among these bacterial systems, *E. coli* is the most used due to low culture cost, high expression levels, easy manipulation, and short turnaround time (Choi *et al.* 2006). Despite its advantages, overexpressing recombinant proteins in *E. coli* often forms insoluble aggregates known as inclusion bodies. In some cases, protein solubility can be improved by selecting appropriate expression vectors, co-expressing with soluble protein or co-expressing with solubility-enhancing proteins. Protein aggregation tends to occur when proteins fail to fold correctly into their three-dimensional structures, resulting in the formation of insoluble protein particles.

The study by Kuri and Goswami, (2024) demonstrated that VP6 forms classical inclusion bodies in *E. coli*, which are misfolded aggregates rather than non-classical inclusion bodies that retain a more native-like structure. This allows the structural protein to undergo harsh solubilisation conditions, such as higher concentrations of urea. The authors reported that the recombinant VP6 was expressed in *E. coli* BL21(DE3) cells and despite different optimization efforts, the protein was still expressed in inclusion bodies rather than in the soluble fraction. To recover VP6 protein from inclusion bodies, the authors used a high concentration of urea (6 M to 8 M), which successfully solubilized and denatured VP6. Whereas with lower urea concentration (less than 6 M) proved ineffective in recovering the VP6 protein under the conditions tested. After solubilization, VP6 was purified using Ni–NTA (His-tag)

affinity chromatography, with an on-column renaturation. Aggregation persisted after refolding, as confirmed by both transmission electron microscopy (TEM) and continued positive Thioflavin T binding, suggesting incomplete refolding by maintaining aggregation/misfolded conformations. This behaviour differs from the non-classical inclusion bodies, which are typically less compact and easier to solubilise due to partial preservation of native structure. Both intrinsic factors, such as amino acid sequence, structural stability, and aggregation-prone hydrophobic regions, and extrinsic factors, such as pH, temperature, ionic strength, and induction conditions, influences the development of aggregation (Wen *et al.* 2023). Exposure of hydrophobic regions during misfolding promotes interactions between protein subunits, causing the protein to aggregate (Fink, 1998). The resulting protein aggregates differ significantly in size, shape, and morphology, depending on the nature of the intermolecular interactions. Therefore, having a deeper understanding of how these aggregates form and their structural characteristics is crucial for effective control and prevention (Mahler *et al*. 2009).

Protein structure directly affects its expression and assembly, particularly in the case of capsid proteins, which ideally need to be expressed in a soluble form to facilitate correct assembly. Protein aggregation presents a significant challenge in biopharmaceutical research and development, particularly in the production of protein-based vaccines. A deeper insight into the mechanisms driving protein aggregation is crucial for developing strategies to manufacture the protein vaccine. For example, when the rotavirus VP6 protein is isolated, it can self-assemble forming either spherical globular or tubular structure depending on the pH. The structures include VLPs and nanotubes (Ready and Sabara, 1987). According to Zhao *et al.* (2011), expressing VP6 in *E. coli* frequently led to the formation of inclusion bodies (insoluble aggregates) that complicated downstream processing, and are time-consuming to handle. Since aggregates require an additional step of denaturation and refolding to recover biologically active protein, chemical denaturants were used during purification to assist in solubilizing the inclusion bodies and help minimize further aggregation. The authors successfully expressed the rotavirus structural protein VP6 in *E. coli* BL21(DE3) strain, resulting in inclusion bodies formation after induction. The researchers solubilized these inclusion bodies using harsher or higher concentrations of urea between 6 M to 8 M to disrupt the protein aggregates. The solubilized protein

was then subjected to purification and refolding processes to recover biologically active VP6.

1.8 Purification of Rotavirus VP6 Protein

Chromatography is a method used to separate, identify, and purify different substances from complex mixtures. It can be applied to both qualitative and quantitative analysis. During purification of a protein, the protein can be separated by various physicochemical properties, including molecular size and shape, net charge, surface hydrophobicity, and specific binding affinity to a stationary phase (Coskun *et al.* 2022). The most used chromatography techniques for purification are ion exchange chromatography (IEX) (separates based on charges), size exclusion chromatography (SEC) (separates based on sizes), and affinity chromatography (which separates based on specific binding interactions such as those between a His-tag and nickel resin, or an enzyme and its ligand (Jungbauer and Hahn, 2009; Barth *et al.* 1998; Kanoh *et al.* 2023). The rotavirus VP6 protein purification most commonly uses affinity chromatography to purify using the His-tagged constructs, which has proven effective for isolating the protein under native conditions (Zhao *et al.* 2011; Bugli *et al.* 2014). This method has been widely applied in studies where VP6 is expressed in *E. coli*, providing high yields of purified protein suitable for structural and diagnostic applications.

Choi *et al.* (2002) successfully expressed VP6 and VP7 in mammalian cells using the Semliki Forest virus (SFV)-based expression system. The VP6 and VP7 genes were cloned into an SFV vector and introduced into the BHK-21 cells via electroporation. Both proteins were successfully expressed in the cytoplasm, as confirmed by immunocytochemistry, Western blot, and electron microscopy. The study demonstrated the feasibility of producing soluble VP6 protein in a eukaryotic environment, providing a suitable platform for further functional and structural studies of rotavirus proteins. The SFV-based system proved to be an effective and safe, as the expressed proteins showed signs of being properly folded and assembled, supporting its potential use for vaccine development or diagnostic applications. Kumar *et al.* (2016) clone the VP6 gene from a bovine rotavirus strain into the pET-32a (+) vector to allow expression of a His-tagged recombinant VP6 (rVP6) protein in *E. coli*. Among the various strains tested, *E. coli* Tuner (DE3) pLysS yielded the highest expression levels following IPTG induction. The expressed rVP6 were localized

primarily in inclusion bodies and they were solubilized using 6 M guanidine hydrochloride. Purification was carried out under denaturing conditions using Ni-NTA affinity chromatography. The protein was refolded on-column by gradually reducing the urea concentration in the wash buffers. The identity and purity of rVP6 were confirmed by SDS-PAGE and Western blotting, indicating successful expression and purification of antigenically active VP6 suitable for downstream diagnostic applications.

Li *et al.* (2019) engineered a fusion protein by genetically linking the rotavirus inner capsid protein VP6 to *Helicobacter pylori* ferritin, with the goal of producing a self-assembling nanoparticle vaccine. The chimeric VP6 ferritin gene was expressed in *E. coli*, where the fusion protein accumulated predominantly as inclusion bodies., The protein was purified using a His-tag resin (Ni-NTA) and further refined by fast protein liquid chromatography (FPLC) to obtain a high-purity preparation. The protein's antigenicity was confirmed by its ability to bind anti-VP6 antibodies, and transmission electron microscopy confirmed the formation of uniform spherical nanoparticles, resembling ferritin nanocages. The authors proved that VP6-ferritin fusion protein can be efficiently expressed in *E. coli*, where it self-assembles into stable nanoparticles that function as effective nano-vaccines capable of inducing strong humoral and mucosal immune responses. These findings provide a solid foundation for the development of an oral rotavirus vaccine. Bugli *et al.* (2014) expressed rotavirus VP6 protein in *E. coli* using three constructs: His-tagged, thioredoxin, and small ubiquitin-like modifier (SUMO) fusions system. The His-tagged VP6 formed insoluble inclusion bodies, while the SUMO fusion greatly improved solubility. After cell lysis and centrifugation, soluble fractions were purified using the affinity chromatography targeting each tag. The SUMO fusion enabled high-yield purification of soluble VP6, which could be enzymatically cleaved to obtain native protein. Purified VP6 self-assembled into virus-like particles and nanotubes, demonstrating proper folding and offering promising nanoconstructs for vaccine development and other applications.

The improvements in recombinant protein expression and purification, particularly through the development of engineered strains like NiCo21(DE3) show significant improvement in addressing limitations associated with traditional BL21(DE3) systems. Robichon *et al*. (2011) played an important role in improving this field by demonstrating that selective strain engineering can effectively reduce non-specific binding and

enhance the yield and purity of His-tagged proteins. The authors work highlighted the importance of combining genetic optimization with optimised purification strategies, such as chitin bead binding, to achieve more consistent and high-quality outcomes. These findings not only strengthen the reliability of protein purification workflows but also set a foundation for future novelties in recombinant protein production.

1.9 Computational Approach for Protein Modelling and Optimisation

Previous studies have demonstrated successful expression of the rotavirus VP6 protein using various expression systems, including *E. coli*, baculovirus-silkworm, and mammalian cells as per Section 1.8, where *E. coli* remains the most commonly used expression system, primarily due to its low cost, high expression levels, and easy manipulation. However, challenges associated with *E. coli*-based expression limited VP6 protein production, primarily due to the frequent formation of inclusion bodies resulting from improper protein folding. This proved challenging when producing biologically active proteins, especially for vaccine development, since the proper structural integrity and solubility of the VP6 protein are crucial for correct assembly into VLPs.

Computational approaches offer solutions to protein challenges by allowing in silico modelling of protein structure, folding pathways, and aggregation-prone regions using methods such as advanced algorithms and molecular dynamics simulations to assist in predicting the structural conformations, identifying mutations that enhance solubility and producing the most improved and efficient expression system (Kuriata *et al.* 2019). These can significantly improve the recombinant protein yield and quality. The tools can also allow the modelling of VP6's structural changes in response to different pH and temperature, which provide useful information during the optimizations and formation of VLPs or nanotubes (Kuri and Goswami, 2024). With complications that occur during protein expression, protein solubility, protein purification, and protein characterization, the addition of computational tools into experimental workflows can help reduce costs and time by increasing the chances of producing a properly folded and functional protein. Some of the most used computational tools are the National Center for Biotechnology Information (NCBI) database, the Jalview tool, the AlphaFold (ColabFold: AlphaFold2 w/ MMseqs2), and the AGGRESCAN3D tool. All these tools are key for obtaining, aligning, and predicting the three-dimensional (3D) structure of proteins, as well as assessing their expression and aggregation potential.

The National Center for Biotechnology Information (NCBI) Reference Sequence (RefSeq) database plays a role in genomic research and annotation by focusing on collecting nucleotide and protein sequences and the database can be accessed through Entrez, BLAST, E-Utilities, and FTP (Pruitt *et al.* 2008). This database is used to provide a high-quality sequence for different organisms such as humans, animals, plants, and microorganisms. Each RefSeq entry has a unique accession format, and it is used to retrieve different sequences. The Jailview tool allows visualization and analysis of multiple protein or nucleotide sequences, aligning them to generate a consensus sequence. (Waterhouse *et al.* 2009). The tool can also predict the variations occurring throughout the alignment of the sequences, and it can also perform functions such as calculating sequence conservation, predicting secondary structures, and identifying motifs.

The AlphaFold (ColabFold: AlphaFold2 w/ MMseqs2) tool is an AI that is used to generate multiple sequence alignments (MSAs) to predict the structure of the protein. The tool was developed by DeepMind and the bioinformatics community, and it uses MMseqs2 (Many-against-Many sequence searching) to generate multiple sequence alignments (MSAs) (Mirdita *et al.* 2022; Yang *et al.* 2023). AlphaFold2 is a good advancement in protein structure prediction, offering near-experimental accuracy. Unlike traditional homology modelling methods, which rely heavily on the availability of closely related template structures, AlphaFold2 can predict accurate protein models even in the absence of homologous templates. During the structure prediction, ColabFold significantly accelerates the process by substituting AlphaFold2's homology search with the 40–60-fold faster MMseqs2 and further improves or speeds up predictions by 90-fold by avoiding or eliminating the need for model recompilation (Mirdita *et al*. 2019). AlphaFold predicts five different structural models for a given protein sequence. These models are ranked based on confidence metrics, and the highest-ranked model is typically selected as the final prediction. AlphaFold assigns a per-residue confidence score called pLDDT (predicted Local Distance Difference Test) to each residue, which reflects the model's confidence in the local positioning of that residue's alpha carbon (Cα). A high pLDDT score (close to 100) indicates that AlphaFold is confident in the predicted position of that residue (Jumper *et al.* 2021).

The AGGRESCAN3D tool is a structure-based database that identifies surface-exposed aggregation-prone regions in proteins by assessing how mutations influence

both the protein stability and aggregation propensity (Carija *et al.* 2019; Kuriata *et al*. 2019). It can be used not only to detect aggregation hotspots but also to guide mutation-based strategies for improving protein solubility. During protein mutation design to reduce aggregation-prone regions, it is crucial to avoid disrupting functional sites, particularly surface regions involved in protein–protein interactions. To preserves the protein functions, both sequence conservation and structural context are essential. AGGRESCAN3D reports residue-level aggregation scores, where positive values (score > 0.000) indicate an increased tendency to aggregate. Residues with higher scores are therefore flagged as aggregation prone. These residues can then be systematically mutated to their substitutions or alternative amino acids, and the mutated sequences are re-evaluated with AGGRESCAN3D to identify variant with reduced aggregation potential while maintaining structural integrity. Importantly, while these substitutions often lead to only slight changes in predicted stability, the temperature-based experimental data indicate that even minor shifts can result in meaningful structural and functional consequences (Kuriata *et al.* 2019; Pujols *et al.* 2018; Gil-Garcia *et al.* 2018). Therefore, it is important to consider both aggregation propensity and stability predictions together when evaluating the biophysical impact of mutations when exploring further in this study.

Parbhoo *et al.* (2016) analysed the rotavirus capsid proteins VP2, VP6, and VP7 sequence conservation and structural roles. The authors concluded that despite the RNA genome mutation rate, the protein remained highly conserved, which is important for maintaining the structural integrity. This was achieved by using computational tools only. The VP2, VP6, and VP7 sequences were retrieved and aligned to generate the consensus sequences and calculate their conservation scores. The VP2, VP6, and VP7 sequences and PDB codes were used to construct their homology models and PyMOL (The PyMOL Molecular Graphics System, Version 0.99rc6 Schrödinger, LLC) was used for structure rendering and visualisation. All these analyses were conducted to gain a better understanding of the protein despite various mutations, aiming to identify regions less affected by sequence variation, with the goal of informing rational vaccine design

1.10 VP6 Protein Structural Conformation and Stability

Different spectroscopy techniques are widely used to detect, identify, and quantify molecular compounds and structural components within a sample. The techniques

include the UV-visible spectrophotometry, which relies on the Beer-Lambert-Bouguer Law to analyse sample absorbance and concentration. Spectrophotometric analysis is important for concentration determination and is mostly used for quantifying DNA, RNA, and proteins (He, 2011; Dadi and Yasir, 2022).

The Far-UV circular dichroism (CD) is used to analyse the secondary structure of the proteins by measuring the difference in absorption between left and right-circularly polarized light in the wavelength range within 180 nm to 250 nm (Kelly *et al.* 2005). This wavelength range reflects electronic transitions of the peptide backbone, which are highly sensitive to conformational changes. The two primary transitions responsible for UV absorption are the n→π* transition, occurring around 210 to 230 nm, and the π→π* transition, which takes place around 180 to 215 nm (Bohnert and Sheen, 1986; Sharma *et al*. 2008). These transitions produce characteristic ellipticity patterns in Far-UV circular dichroism spectra that relate with specific secondary structures. For example, the alpha-helical proteins exhibit a distinct double minimum at 208 nm (π→π*) and 222 nm (n→π*), and a strong maximum near 191–193 nm. In contrast, beta-sheet structures are characterized by a minimum at 215 nm (n→π*) and a maximum at 198 nm (π→π*) (Kelly *et al.* 2005; Wei *et al*. 2014). Random coils display spectral patterns with less defined features, opposite to those of alpha-helices and beta-sheets. These spectral features assist in understanding the structural changes of the protein secondary elements. Under stress conditions like heat or exposure to chemical denaturants, proteins can unfold or aggregate, and Far-UV circular dichroism can detect these transitions (protein folding, conformational stability, and dynamic structural transitions) by monitoring shifts in ellipticity at characteristic wavelengths.

Attenuated Total Reflectance Fourier Transform Infrared (ATR-FTIR) spectroscopy is a powerful technique for analysing the secondary structure of proteins (Srour *et al*. 2017; Sandt, 2024; Tintor *et al.* 2024). This method is invaluable for investigating protein folding, conformational changes, and molecular interactions. Furthermore, ATR-FTIR can be effectively combined with other techniques, such as Far-UV circular dichroism spectroscopy, to provide a more comprehensive understanding of protein structure. The ATR-FTIR focuses on the determination of the protein secondary structures across a broader range of concentrations, and it can predict the information

about the chemical composition of a sample. Each molecular structure produces a specific infrared spectrum, more like a fingerprint by displacing its absorbance and transmission patterns, making infrared spectroscopy a valuable tool for various analytical applications. ATR-FTIR is particularly advantageous for infrared analysis because it allows direct measurement of samples with minimal preparation, enabling detection of characteristic amide I and II bands associated with protein backbone conformations. Proteins can be characterized by nine FTIR absorption bands, known as amide A, B, and amide I through VII. Among these, the most informative for protein analysis are the amide I, amide II, and amide A bands (Krimm and Bandekar, 1986; Goormaghtigh *et al.* 2006). The amide I band (1700 cm^{-1}–1600 cm^{-1}) is the most sensitive region for detecting protein secondary structure, primarily arising from C=O stretching vibrations of the peptide backbone, which contribute approximately 80% of the signal and it can be used to distinguish between alpha-helices, beta-sheets, turns, and unordered structures. In contrast, the amide II band (1510 cm^{-1}–1580 cm^{-1}), originating from in-plane N–H bending (accounting for 40–60% of the potential energy) and C–N stretching vibrations (18–40%) is less responsive to conformational changes compared to the amide I band. (Krimm and Bandekar, 1986; Yang *et al.* 2015; Sadat and Joye, 2020). Amide A band (3310 cm^{-1} –3270 cm^{-1}) mainly corresponds to the N–H stretching vibration of the peptide backbone in proteins. Shifts in the Amide A band can indicate changes in the protein's secondary structure or its environment, such as the formation or disruption of hydrogen bonds (Barth, 2007). FTIR spectroscopy is a convenient technique that requires minimal time and is not restricted by molecular weight. The ATR-FTIR spectroscopy is also known for detecting molecular aggregation by capturing changes in specific vibrational bands, such as shifts in amide or hydroxyl stretching regions, that reflect altered intermolecular interactions or structural changes/conformations (Tintor *et al.* 2024).

Intrinsic fluorescence spectroscopy is commonly used to investigate the tertiary structure of proteins by analysing the emission spectra of the tryptophan residues and it is a very sensitive and cost-effective method for studying the protein–ligand interactions, structural dynamics, and environmental changes (Lakowicz, 2006; Rossi and Tylor, 2011; Ghisaidoobe and Chung, 2014). For the 1QHD code for the VP6 protein, the alpha helix domain contains three tryptophan's (position 59,139 and 367) and the beta sheet domain contains two tryptophan's (position 181 and 247) (Mathieu

et al. 2001) (Figure 1.4). Tryptophan (Trp) plays a significant role in protein UV absorbance, particularly at 280 nm, and it is the main contributor of the intrinsic fluorescence typically between 320 nm and 360 nm (Ghisaidoobe and Chung, 2014; Teale and Weber, 1957). Even though Trp is the most sensitive, other aromatic residues such as tyrosine and phenylalanine can also contribute to the fluorescence signal, although to a lesser extent due to their lower quantum yields. The exact emission maximum depends on the local environment of the Trp residue, including its solvent accessibility and the structural state of the protein (Lakowicz, 2006; Hixon and Reshetnyak, 2009; Chen and Barkley,1998).

The fluorescence behaviour of Trp is often monitored to study the protein's tertiary structure, as shifts in the emission maximum and variations in intensity can reflect changes in the residue's surrounding environment (Vivian and Callis, 2001; Dos Santos Rodrigues *et al.* 2023.). A shift to a shorter wavelength is known as a blue shift, which indicates a transition to a more hydrophobic (nonpolar) environment. In contrast, a red shift, or a longer wavelength, suggests increased polarity or solvent exposure. During protein unfolding, for example, Trp residues become more exposed to aqueous surroundings, leading to a red shift in emission as the environment around them becomes more polar (Brack *et al.* 2024). Thermal denaturation also affects Trp fluorescence, as an increase in temperature can disrupt the protein's tertiary structure, leading to increased solvent exposure of Trp residues and a corresponding red shift in emission. In addition, fluorescence quenching can occur due to various factors such as nearby quenching groups (e.g. disulfide bonds, protonated histidine, or even solvent molecules) or through dynamic quenching mechanisms, which further influence the intensity of the fluorescence signal (Gooran and Kopra, 2024). Monitoring these changes under thermal stress provides valuable insight into protein stability and folding dynamics.

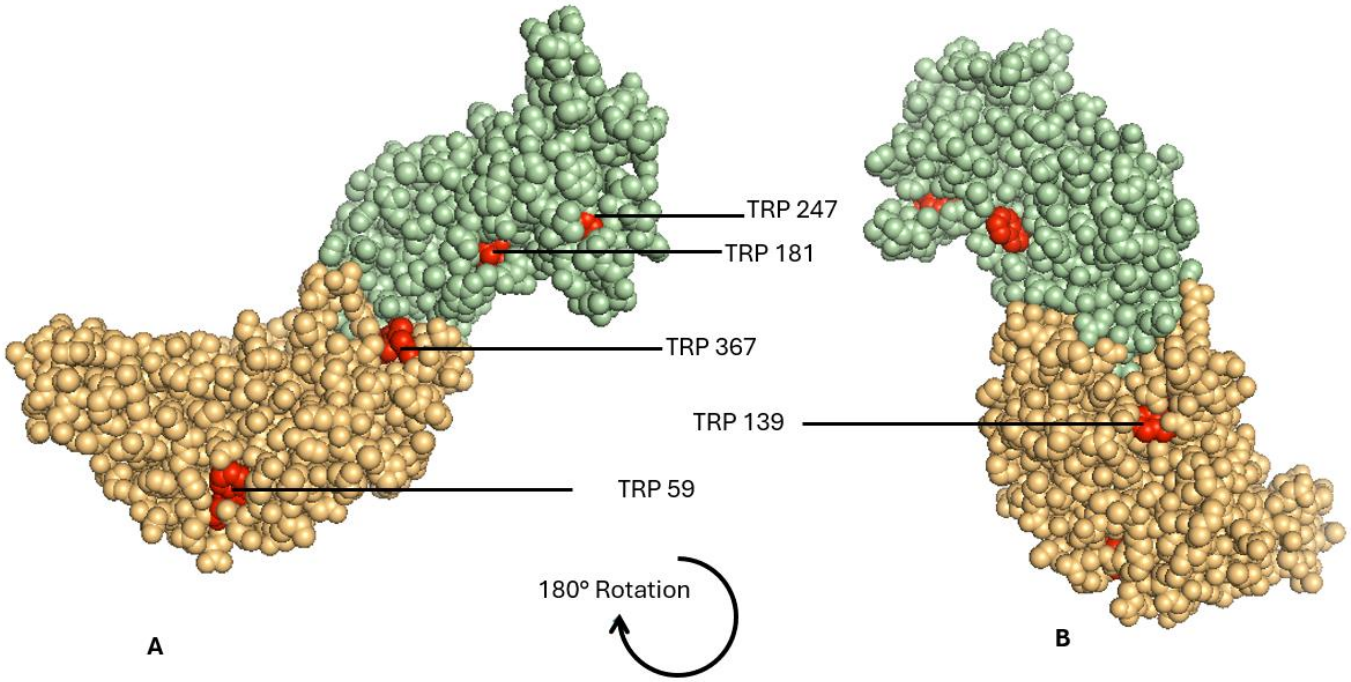


Figure 1.4: VP6 Protein Tryptophan Positions

Shows assembly of VP6 structure with alpha helical domain represented by the yellow region and the beta sheet domain represented by the green region. **(A)** Showed the original view of the tryptophan residues positions highlighted in red colour and **(B)** showed the same structure rotated 180° to view the opposite side. The structure was Generated using PyMOL (The PyMOL Molecular Graphics System, Version 3.0 Schrödinger, LLC). The VP6 PDB structure code: 1QHD (Mathieu *et al.* 2001).

2 Research Aim and Objectives

To address the challenge of poor solubility in the VP6 protein, the study aimed to optimize a consensus sequence of VP6 with enhanced solubility while remaining within the constraints of natural sequence conservation. The optimized protein was intended for efficient expression in *E. coli* and purification using Immobilized Metal Affinity Chromatography (IMAC) to obtain a high yield of pure, soluble VP6 protein suitable for use in rotavirus vaccine development. The study consisted of four objectives. First, the VP6 sequences were compiled from databases (from 2017 to 2024), and they were used to generate a conserved consensus sequence. Second, bioinformatic tools were used to identify regions on the consensus sequence that are prone to aggregation, and alternative amino acid substitutions were determined accordingly to improve solubility while preserving conservation, thereby producing an optimized VP6 sequence. Third, the optimised genes were codon optimised, synthesized, and cloned into a pET-15b plasmid vector for expression in *E. coli* strains BL21 (DE3) and NiCo21 (DE3). Different IPTG concentrations, incubation times, and temperatures for optimal expression and solubility were tested. The expressed proteins were purified using IMAC. Fourth, the purified VP6 proteins were characterised by assessing their structural integrity under native conditions, as well as in the presence of denaturants and thermal stress, including analysis of their secondary, tertiary, and quaternary structures.

3 Methodology

3.1 Materials

The experimental materials used in this study included the VP6 encoding genes (the optimised VP6 consensus wild-type sequence and VP6 mutant sequence), these were codon-optimised and synthesized then inserted into the pET15b vector at the NdeI and BamHI restriction sites. Synthesis and cloning were performed by Inqaba Biotechnical Industries (Pty) Ltd (Pretoria, South Africa) which outsourced and used Biomatik Corporation (Cambridge, Ontario, Canada). Ampicillin and isopropyl-β-D-1-thiogalactopyranoside (IPTG) were purchased from Melford (Ipswich, United Kingdom). *Escherichia coli* (*E. coli*) strains JM109 cells, NiCo21 (DE3) cells, and BL21 (DE3) cells were purchased from New England Biolabs Inc. (Ipswich, United Kingdom). The pre-packed HisTrap™ HP columns and HisTrap™ Desalting columns were purchased from Cytiva (Massachusetts, United States). Ultra-pure urea was purchased from Melford (Ipswich, United Kingdom). Imidazole was purchased from Sigma-Aldrich (United States of America). All other chemicals and reagents were of analytical grade.

3.2 Protein Consensus Sequence Identification and Engineering

The National Center for Biotechnology Information (NCBI) database (https://www.ncbi.nlm.nih.gov/) was accessed on January 18, 2024, to retrieve all Rotavirus A (RVA) VP6 protein sequences starting from the year 2017 that contained the full length sequence of 397 amino acids using the search string Rotavirus A and (Host: human) and VP6 and 397 [sequence length] from year 2017/01/01 to year 2024/01/18 (National Center for Biotechnology Information, 1988). The acquired VP6 sequences were saved in FASTA format. The sequences were multi-aligned using the Jalview tool (http://www.jalview.org/) accessed on January 18, 2024 (Waterhouse *et al.* 2009), to generate the consensus sequence and visualise regions of variability present, and examine alternative amino acids that may influence the structure or function of the protein. AlphaFold (ColabFold: AlphaFold2 w/ MMseqs2) was accessed on January 19, 2024 (Mirdita *et al.* 2022) to generate consensus sequence structure, hence the pdb file. This was achieved by inserting the consensus sequence on the AlphaFold ColabFold query sequence search engine and retrieve the VP6 protein predicted structures. AlphaFold2 was selected instead of homology modelling or

mutating the 1QHD structure due to its high accuracy in predicting the 3D structure of the protein, even in the absence of closely related homologs. It is particularly effective at reflecting the structural impact of sequence variations (Mirdita *et al.* 2019; Steinegger and Soding, 2017). ColabFold, a user-friendly implementation of AlphaFold2, enhances prediction speed by replacing the standard homology search with the significantly faster MMseqs2 algorithm (Mirdita *et al.* 2022), which offers a 40–60-fold improvement in single-sequence predictions. Additionally, it accelerates batch processing by approximately 90-fold through optimizations such as avoiding recompilation and introducing early stopping, making it a good option for producing the best model. The database was found to be the best when creating the pdb files and structures, which were unique identifiers for a 3D model of the VP6 rotavirus structure, offering valuable insights into the protein's structure.

The consensus sequence was submitted to the AlphaFold prediction engine, which produced five structural models. Among these, the model that most closely resembled the VP6 protein structure was selected as the best fit. Each model was evaluated using the Local Distance Difference Test (lDDT-Cα), which compares Cα atom positions between predicted and known structures, with lower differences indicating higher accuracy and each model comes from one of the five different AlphaFold2 checkpoints ensuring diversity in predictions, for single chain (monomeric) predictions: all five models are ranked according to their average predicted Local Distance Difference Test score (pLDDT) which is a per-residue confidence score (0–100) estimating local structure reliability and out of five model the model with the highest mean pLDDT is deemed the most confident and labelled as rank_001 (Mirdita *et al.* 2022; Jumper *et al.* 2021). Based on this metric, models were examined for overall and regional confidence, with values below 50% as low-confidence and values around or above 90% as high-confidence regions. The selected model in a pdb file format was further analysed by identifying amino acid residues that contained different variations and are prone to aggregation using the AGGRESCAN3D tool.

The AGGRESCAN3D tool (http://biocomp.chem.uw.edu.pl/A3D2/) was accessed on January 28, 2024 (Conchillo-Solé *et al.* 2007), to predict the "hot spots" of aggregation for the VP6 consensus sequence from the pdb file generated by AlphaFold. To reduce aggregation, aggregation-prone residues on the VP6 consensus sequence were changed to their ideal substitutions identified. The process prioritized amino acid

substitutions that sustained similar biological characteristics (such as size, charge, and hydrophobicity) to ensure the protein's structural integrity and functionality were unaffected, ultimately aiming to reduce aggregation tendencies without introducing excessive variability that could potentially impair the protein's activity. The mutation procedure involved mutating one amino acid residue at a time in the VP6 consensus sequence. Following each mutation, the AGGRESCAN3D tool was used to calculate the aggregation score of the modified residue. The most suitable amino acids substitution was chosen based on the lowest aggregation score. The Discovery Studio Visualizer v21. 1.0.20298 (https://discover.3ds.com/) accessed on January 29, 2024, was used to visualize and incorporate the amino acid residues with the best aggregation scores into the mutated sequence (Baroroh *et al.* 2023). This technique was repeated for six amino acids, changing one amino acid at a time, to optimize the sequence with the lowest aggregation propensity but still within variation of the consensus sequence obtained.

After the mutation of all six amino acids, the VP6 wild-type sequence and VP6 mutant sequence were analysed using the Expasy ProtParam tool (https://web.expasy.org/protparam/) accessed on January 30, 2024 to assess the properties such as molecular weight, instability index, and amino acid composition (Gasteiger *et al.* 2005). The VP6 wild-type sequence and VP6 mutant sequence were uploaded into the Protparam tool search engine, which calculated various physicochemical properties, including molecular weight, theoretical isoelectric point (pI), amino acid and atomic compositions, extinction coefficient, instability index, aliphatic index, and the grand average of hydropathicity. Then VP6 wild-type sequence and VP6 mutant sequence were reverse translated into nucleotide sequences using the NovoPro (Expoptimizer) tool (https://www.novoprolabs.com/tools/codon-optimization) accessed on January 31, 2024 (Yin *et al.* 2023). To improve gene expression in *E. coli*, codon optimization was performed by modifying the DNA sequence to align with *E. coli's* preferred codon usage. Following the optimization of the genes, Inqaba Biotechnical Industries (Pty) Ltd (Pretoria, South Africa) was used as a supplier for the synthesis of the gene encoding sequence (VP6 encoding gene) and its cloning into the NdeI and BamHI restriction sites on the multiple cloning site (MCS) of the pET-15b expression plasmid (https://www.genscript.com/expression-vector-selection-guide.html). The selection of the NdeI and BamHI restriction sites on

the pET-15b plasmid allowed for the inclusion of the hexahistidine (His6) tag sequence positioned at the N-terminal, facilitating protein purification steps post-synthesis. The tag, composed of six consecutive histidine residues, facilitates protein purification via immobilized metal ion affinity chromatography (IMAC) by providing a selective binding site for metal ions. The plasmid contained an ampicillin resistance (Amp) gene, which serves as a selectable marker, allowing bacteria to grow in media containing ampicillin.

3.3 Bacterial cells Transformation and Cells selection

3.3.1 *Escherichia coli* JM109 Cells Competency

The non-competent JM109 cells were grown overnight at 37°C with shaking at 200 rpm for 12-16 hours in a 100 ml Super Optimal Broth (SOB) containing 1.0% Bacto-tryptone, 0.5% Bacto-yeast extract, 10.0 mM NaCl, 2.5 mM KCl, and 10.0 mM $MgCl_2$. The 2 ml overnight culture was subcultured into a fresh 100 ml SOB and incubated at 37°C until the culture reached an OD_{600}, measured using the Implen P330 Nanophotometer (München, Germany). Once the desired optical density was reached, the cells were harvested by centrifugation at 2264 × *g* for 10 minutes at 4°C. The cell pellets were resuspended in 12.5 ml of transformation buffer (TFB) solution containing 10 mM MOPS, pH 6.2, with 100 mM KCl, 45 mM $MnCl_2{\cdot}4H_2O$, 10 mM $CaCl_2{\cdot}2H_2O$, and 3 mM $HaCoCl_3$ and then incubated on ice for 15 minutes. After incubation, the cells were pelleted again by centrifugation at 2264 × *g* for 10 minutes at 4°C. Subsequently, the cell pellet was resuspended into the 2.5 ml TFB solution, and 3.5% (84 µl) of Dimethyl sulfoxide (DMSO) was added to the mixture. The mixture was then chilled on ice for 5 minutes. Afterward, 84 µl of 75 mM Dithiothreitol (DTT) was added, followed by another 10 minutes of incubation on ice. An additional 84 µl of DMSO was added to the mixture, which was again chilled on ice for 5 minutes. Finally, 21 µl aliquots of the prepared competent cells were pipetted into 1.5 ml microcentrifuge tubes and stored at -80°C for future transformation procedures.

3.3.2 *Escherichia coli* BL21(DE3) and NiCo21(DE3) Competency

The non-competent *E. coli* BL21(DE3) and NiCo21(DE3) cells were grown in a 100 ml Luria-Bertani broth containing 1.0% (w/v) tryptone, 0.5% (w/v) yeast extract, and 0.5% (w/v) sodium chloride at 37°C with shaking at 220 rpm for 12-16 hours. From the overnight cultures, 2 ml aliquots were subcultured into 100 mL of freshly prepared sterile LB broth and incubated until they reached an optical density (OD_{600}) suitable

for harvest. Cells were then harvested by centrifugation at 2264 × *g* for 10 minutes at 4°C. The resulting cell pellets were resuspended in 10 ml of chilled 0.1 M calcium chloride ($CaCl_2$) and kept on ice for 30 minutes. The cells were again pelleted by centrifugation at 2264 × *g* for 10 minutes at 4°C, and the final pellets were resuspended in 1 mL of 0.1 M $CaCl_2$ + 15% glycerol. Competent cells were then aliquoted into 50 µl in 1.5 ml microcentrifuge tubes and stored at -80°C for future use.

3.3.3 Bacterial Cells Transformation

Competent *E. coli* strains BL21(DE3), JM109, and NiCo21(DE3) were first thawed on ice for 10 minutes. Each cell suspension was then mixed with 5 µg of recombinant plasmid DNA (pET15b encoding Rotavirus A viral protein 6 (VP6) wild-type and mutant). The mixtures were then chilled on ice for 30 minutes. The cells were subsequently heat-shocked at 42°C for 60 seconds using a heating block and then immediately cooled on ice for 2 minutes. After the heat shock, Super Optimal Broth (SOC) media was added to the mixtures and incubated at 37°C with shaking at 220 rpm for 70 minutes to allow for recovery. The transformed cells were plated into, Luria-Bertani agar (1.0% (w/v) tryptone, 0.5% (w/v) yeast extract, 1.0% (w/v) sodium chloride, and 1.5% (w/v) agar) with 0.1 mg/ml of ampicillin to allow for selected growth of the *E. coli* BL21(DE3), JM109, and NiCo21(DE3) cells with recombinant plasmid pET15b encoding Rotavirus A viral protein 6 (VP6) wild-type protein and VP6 mutant proteins. The spread plate method was used for plating, and the plates were incubated at 37°C for 12–16 hours to allow colony growth.

3.4 Glycerol Stocks

A single transformed colony from each of the JM109, BL21(DE3) and NiCo21(DE3) strains for the VP6 wild-type protein and VP6 mutant protein was selected and grown overnight in a 100 ml of freshly prepared sterile Luria-Bertani (LB) broth supplemented with 0.1 mg/mL ampicillin at 37°C with shaking at 220 rpm. Ampicillin was used because of the resistance conferred by the plasmid pET-15b. Following overnight growth, 2 ml of the cultures were inoculated into 100 ml of fresh sterile LB broth containing 0.1 mg/ml of ampicillin and incubated at 37°C with shaking at 220 rpm. The OD_{600} was monitored at time intervals until the cells reached the mid-log phase, approximately OD_{600} 0.6 to prepare the glycerol stock. For glycerol stock preparation,

equal volumes of cell cultures and sterile 30% glycerol stock were mixed in a 1:1 ratio, and the stock cells were then stored at -80°C.

3.5 DNA Extraction

The DNA extraction method began with the selection and growth of single colonies of JM109 strain transformed with either the VP6 wild-type encoding plasmid or VP6 mutant encoding plasmid. These colonies were cultured in a sterile LB Broth supplemented with 0.1 mg/ml of ampicillin at 37°C with shaking at 220 rpm for 12-16 hours.

Following overnight growth, 2 ml of the bacterial cultures were subcultured into a fresh sterile SOB medium supplemented with 0.1 mg/ml of ampicillin. The cultures were incubated at 37°C with shaking at 220 rpm until the OD_{600} between 0.4 to 0.6 was reached, indicating cells were in the mid-log phase suitable for DNA extraction (Zymo Research. 2021). DNA extraction was carried out according to the protocol outlined in the ZymoPURE Plasmid Miniprep Kit (https://zymoresearch.eu/pages/zymopure) accessed April 15, 2024 (Zymo Research, 2021). A 5 ml volume of the bacterial cultures was pelleted by centrifugation at maximum speed for 20 seconds. The supernatant was discarded, and the pellet was resuspended in 250 µl of ZymoPure P1 buffer. Then, 250 µl of ZymoPure P2 (pre-warmed to 37 °C) was added, and the tubes were gently inverted 8 to 10 times before incubating at room temperature for 3 minutes. Next, 250 µl of ZymoPure P3 was added, mixed by gentle inversion, and the mixture was centrifuged at 16000 × *g* for 5 minutes. After centrifugation, 600 µl of the cleared lysate was transferred to a new clean microcentrifuge tube and combined with 260 µl of ZymoPure binding buffer. The resulting solution was vortexed for 15 seconds and then loaded onto a Zymo-Spin II-PX column. After a minute incubation at room temperature, the column was centrifuged at 11000 × *g* for 1 minute and then the flow-through was discarded.

To purify the plasmid DNA, the spin column was subjected to a series of wash steps. Initially, 800 µl of ZymoPure Wash 1, 800 µl of ZymoPure Wash 2, followed by 200 µl of ZymoPure Wash 2 consecutively. After each wash, the column was centrifuged at 11000 × *g* for 1 minute and the flow-through was discarded. An additional centrifugation step was performed for 1 minute at 11000 × *g* to remove any remaining wash buffer residual. For elution, 25 µl of ZymoPure elution buffer (pre-warmed to

50°C) was added to the column. The column was incubated at room temperature for 2 minutes to allow the buffer to saturate the membrane before centrifugation at 11000 × *g* for 1 minute to collect the purified DNA. The concentration and purity of the extracted DNA was analysed using the Ezdrop1000c Micro-Volume Spectrophotometer (Blue-Ray Biotech, Taiwan), measuring both absorbance and concentration.

DNA extraction was successfully conducted using the transformed JM109 which focused on two distinct DNA samples: the VP6 wild-type protein and the VP6 mutant protein. For the VP6 wild-type protein, the extraction process yielded high-quality genomic DNA with a notable 260/280 nm ratio of 1.78. Similarly, for the VP6 mutant protein, the extraction process yielded high-quality genomic DNA, with an even higher 260/280 nm ratio of 1.93. This ratio signifies minimal protein contamination, indicating the purity of the extracted DNA. The DNA extraction was performed to amplify plasmid DNA in the cell system, and the DNA was further used to transform the BL21 (DE3) and NiCo21 (DE3) and further use.

3.6 Bacterial Growth Curves and VP6 Protein Expression

3.6.1 Bacterial Growth Curve

To analyse the growth curve, glycerol stock of *E. coli* strains BL21 (DE3) and NiCo21 (DE3) containing plasmid encoding VP6 wild-type protein and VP6 mutant protein were grown using 100 ml 2X YT media (1.5% (w/v) tryptone, 1% (w/v) yeast extract, 0.5% (w/v) sodium chloride with 0.1 mg/ml of ampicillin at 37°C with shacking 220 rpm for 12-16 hours. From the overnight-grown cells, 2ml of each culture was subcultured into fresh 2X YT media. Their OD_{600} was then measured using an Implen P330 Nanophotometer (München, Germany). The procedures facilitated the construction of the growth curves using the Microsoft Excel 2016 program, enabling the precise plotting of data points.

3.6.2 VP6 Protein Expression

Glycerol stock of *E. coli* strains BL21 (DE3) and NiCo21 (DE3) containing plasmid encoding VP6 wild-type protein and VP6 mutant protein were grown in 100 ml 2X YT media (supplemented with 0.1 mg/ml antibiotic ampicillin) at 37°C 220 rpm for 12-16 hours. From the overnight culture, 2 ml of each culture was subcultured into freshly made 100 ml of 2X YT media with 0.1 mg/ml antibiotic ampicillin and incubated at two

different temperatures, 37°C and 20°C, under constant shaking at 220 rpm. Once the cultures reached the mid-log phase, protein overexpression was initiated by adding different Isopropyl β-D-1-thiogalactopyranoside (IPTG) concentrations: BL21 (DE3) cultures were induced using 0.1 mM and 1,0 mM IPTG, while NiCo21 (DE3) cultures were induced with 0.04mM and 0.4m IPTG. These IPTG concentrations for BL21 (DE3) and NiCo21 (DE3) were chosen to optimise protein expression while minimising potential cellular stress and maximising protein quality (https://www.neb.com: New England Biolabs). After induction, samples were collected at various time intervals (4 hours, 5 hours, 6 hours, 7 hours, 8 hours, and 12 hours). The cells were collected by centrifugation at 16000 × *g* for 30 minutes at 4°C. The resulting pellets were resuspended in 500µl deionized water (dH_2O). The samples were ultrasonicated using an Q700 ultrasonic liquid processor (QSonica, 53 Church Hill Rd. Newtown, CT, USA) set at 35% amplitude. Each sample underwent three cycles of sonication, consisting of 30 seconds of pulsed sonication (1-second on and 1-second off) per cycle. A 10-minute break on ice was applied between cycles, and all samples were sonicated on ice to ensure efficient cells disruption. After sonication, the samples were further chilled on ice for 10 minutes, and then they were centrifuged at 16000 × *g* for 30 minutes at 4°C. The pellet and supernatant fractions were separated, with pellets resuspended in 500µl dH_2O. Both fractions were analysed using a 12% SDS-PAGE gel as described in section 3.10.1. The resulting SDS-PAGE gels were imaged using a BioRad Gel Dock Universal Hood (Hercules, United States of America).

3.7 Isolation of the VP6 Protein Inclusion Bodies

The isolation of inclusion bodies for VP6 wild-type protein and VP6 mutant protein was carried out following the methodology outlined by Russell and Gildenhuys, (2018), which involves sequential wash steps to effectively isolate the inclusion bodies from other cellular components. The method was used as an additional step to help remove contaminants before purification from trial 4 to trial 5 (VP6 wild-type protein) and trial 1 to trial 3 (VP6 mutant protein), section 3.9.1.

After sonication, the cell lysates were collected by centrifugation at 16000 × *g* for 30 minutes. The pellets were resuspended in 50 ml of Wash Buffer A containing 50 mM Tris-HCl (pH 7.8), and 1% Triton X-100, and they were incubated at room temperature for 30 minutes while shaking to assist in disrupting the inclusion bodies. The pellets

were subjected to two more washes using the same buffer (Wash Buffer A). The buffer contained Triton X-100, which, in addition to producing highly purified inclusion bodies, also assists in cleaning inclusion bodies by removing cell walls and outer membrane fragments (Singh *et al.* 2015; Patra *et al.* 2000). The first washes were followed by the second wash, which contained 50 ml of Wash Buffer B (50 mM Tris-HCl, pH 7.8). The pellets were washed twice with Wash Buffer B to remove any traces of the cell walls, contaminations, and the 1% Triton X-100 from the proteins, since it can interfere with the solubilisation step by either aggregating or denaturing protein (Palmer and Wingfield, 2012). The pellets were further used for solubilisation and purification.

3.8 VP6 Protein Solubilisation

When expressing the VP6 recombinant protein in bacterial systems, inclusion body formation is common. These inclusion bodies are classical in nature and differ from non-classical inclusion bodies, particularly in their structural integrity and the functional state of the encapsulated protein. (Kuri and Goswami, 2024). Sometimes it can be difficult to recover protein from inclusion bodies since the protein is often aggregated and in a non-functional form (Singh *et al.* 2015). Proteins were solubilised using urea as the denaturant. A 10 M urea stock solution was prepared as per Pace, (1986), and the concentration was measured using the Atago R5000 refractometer (Tokyo, Japan), and the final concentration was predicted multiple times from June 10, 2024 using the Sosnick database (https://sosnick.uchicago.edu/gdmcl.html), a denaturant concentration calculator (Sosnick, 2023). Urea was used alongside a freeze-thaw method to solubilise the protein Qi *et al.* 2015.

The *E. coli* strain NiCo21 (DE3), overexpressing either VP6 wild-type protein or VP6 mutant protein as described in section 3.6.2, were harvested 12 hours post-induction with 0.04 mM IPTG. Cells were collected by centrifugation at 16000 × *g* for 30 minutes at 4°C. The inclusion bodies obtained after washing were resuspended in Lysis buffer containing 50 mM Tris-HCl, 150 mM NaCl, adjusted to pH 7.5 or 8.0. The mixtures were centrifuged again at 16000 × *g* for 30 minutes at 4°C to pellet the insoluble material. For solubilisation, the resulting pellets were resuspended in 500µl of the following buffers: Solubilisation buffer A (50 mM Tris-HCl, without Urea) and Solubilisation buffer B containing 50 mM Tris-HCl, with 2 M, 5 M, and 7 M urea, adjusted to pH 7.5 or pH 8.0. A freeze-thaw cycle was used as per Qi *et al.* 2015, wherein lysates were frozen overnight at -20°C and then thawed at room temperature

for 1 hour. After thawing, the lysates were centrifuged at 16000 × *g* for 30 minutes at 4°C to separate the soluble (supernatant) and insoluble (pellet) fractions. These fractions were analysed by SDS-PAGE, as described in section 3.10.1.

3.9 VP6 Protein Purification

3.9.1 Affinity Chromatography

Immobilized Metal Affinity Chromatography (IMAC) is a highly efficient method for purifying recombinant proteins tagged with a histidine (His) tag. The His-tag has a high binding affinity for metal ions (typically nickel or cobalt) immobilized to the surface of the chromatography resin (Kanoh *at al.* 2023). This interaction facilitates selective binding and isolation of the target protein, such as Rotavirus VP6, by effectively separating it from non-specific proteins and other impurities. The procedure usually starts with the HisTrap column, which selectively binds and concentrates the His-tagged protein while washing out non-target proteins and contaminants. Elution of the VP6 protein is achieved by disrupting the His-tag-metal ion interaction, resulting in the recovery of a highly purified protein. IMAC has great specificity, yield, and purity, making it appropriate for protein-based research applications. VP6 Proteins were purified using an IMAC technique on a Cytiva ATKA chromatography system (GE Healthcare Bioscience, Uppsala, Sweden). The Cytiva 5 ml His Trap column was connected to a Cytiva ATKA purification system; the column was equilibrated using an Equilibration buffer at a flow rate of 2 ml/min unless stated otherwise, and the same flow rate was used for a linear elution gradient. Different trials (Table 3.1) aimed to improve the protocol, protein purity, and yield while removing contaminants/impurities effectively were implemented. The solubilized VP6 proteins (supernatant), obtained as per section 3.8, were loaded into the Bio-Rad ATKA Super-loop and then the proteins were injected into a pre-equilibrated HisTrap column, allowing efficient binding to the column, facilitating subsequent purification steps through affinity chromatography using His-tag interaction.

VP6 wild-type Trial 1: The solubilized protein was injected into a pre-equilibrated HisTrap column with Equilibration buffer A (0 mM imidazole). The column was subsequently washed using Wash buffer C at a flow rate of 2 ml/min for 5-column volumes.

Table 3.1: VP6 Protein Purification Overview

Shows an overview of the VP6 wild-type protein and VP6 mutant protein purification using IMAC with different optimized multiple trials to improve yield, purity, and the removal of contaminants. Solubilised proteins were injected into a pre-equilibrated HisTrap column and purified through washes and elution with different imidazole concentrations. The following table summarizes the conditions from each trial, detailing the methods used to achieve optimal protein purification for both VP6 wild-type protein and VP6 mutated protein.

VP6 Wild-type Protein							
Trials	**pH**	**NaCl (mM)**	**Imidazole (mM)**			**Urea (M)**	**Tween-20 (%)**
			Equilibration Buffer	**Wash Buffer**	**Elution Buffer**		
01	8.0	300	0	150	500	5	
02	8.0	300	5	150	500	5	
03	7.5	150	50	150	600	5	
04	7.5	150	80	80	600	7	0.5
05	7.5	150	80	80	600	7	0.5 (300 NaCl)
VP6 Mutant Protein							
01	7.5	150	80	150	600	7	0.5
02	7.5	150	80	150	600	7	0.5
03	7.5	150	80	150	600	7	0.5 (300 NaCl)

Elution was performed using a linear imidazole gradient, gradually increasing the concentration from 150 mM (Wash buffer C) to 500 mM (Elution buffer A) for 20-column volumes. Following elution, the column was further washed with 2-column volumes of 100% buffer B (Elution buffer A). Fractions were collected throughout the process. The elution fractions (2 ml each) were collected during the elution gradient from F1 to F18 and analysed using SDS-PAGE.

Trial 2: The solubilised protein was injected into a HisTrap column equilibrated with Equilibration buffer A containing 5 mM imidazole. The column was then washed with 5-column volumes of Wash buffer C. Elution was performed using a 20-column volume gradient from 150 mM (Wash buffer C) to 500 mM (Elution buffer A). Afterwards, the column was washed with 100% buffer B (Elution buffer A) for 2-column volumes. The fractions (2 ml each) were collected during the elution gradient from F1 to F22 for further analysis.

Trial 3: Solubilized protein was injected into a pre-equilibrated HisTrap column with Equilibration buffer B, where NaCl concentration was decreased from 300 mM to 150 mM. The column was washed with 5-column volumes of Equilibration buffer B, followed by a 10-column volumes gradient wash increasing the imidazole concentration from 50 mM to 150 mM imidazole (Wash buffer C). Protein elution was carried out using a 20-column volume gradient from 150 mM imidazole (Wash buffer C) to 600 mM (Elution buffer B). A final wash with 100% buffer B (Elution Buffer B) was performed for 5-column volumes. The elution fractions (2 ml each) were collected during the elution gradient from F1 to F38 for further analysis.

Trial 4: Purification was optimized by analysing the solubilisation yield with urea concentrations from 5 M to 7 M, the protein's binding capacity, and the contaminants eluting during the purification process. The solubilized protein, as per 3.7 and 3.8 sections was injected into a pre-equilibrated HisTrap column with Equilibration buffer C. The column was washed with 20-column volumes of Equilibration buffer C to remove non-specifically bound proteins. Elution of the His-tagged proteins was performed over 40-column volumes using a gradient that increased imidazole from 80 mM (Equilibration buffer C) to 600 mM (Elution buffer B). This was followed by an additional wash step using 10-column volume of 100% buffer B (Elution buffer B).

Throughout the gradient, 2 ml fractions were collected, specifically from fractions (F4 to F41) for further analysis.

Trial 5: Purification focused on removing contaminants during the elution process. Solubilized protein was injected into a pre-equilibrated HisTrap column with Equilibration buffer C. To enhance the removal of non-specific bound proteins, the column was washed with a 10-column volume of Wash buffer E containing 0.5% Tween-20, NaCl from 150 mM to 300 mM. Han *et al.* 2019 mentioned that a higher concentration of NaCl helps to effectively disrupt non-specific interactions, thereby enhancing target protein purity. Following the detergent wash, the column was washed with 20-column volumes of Equilibration buffer C to remove Tween-20 traces. The bound proteins were then eluted using an imidazole gradient from 80 mM (Equilibration buffer C) to 600 mM (Elution buffer B) over 40-column volumes. This was then followed by a final wash using 10-column volumes of 100% Buffer B (Elution buffer B). Throughout the elution profile, 2 ml fractions were collected, specifically from fractions F11 to F70, for further analysis.

VP6 Mutant Protein Trial 1: Purification was optimized by analysing the solubilisation yield with urea concentrations from 5 M to 7 M, the protein's binding capacity, and the contaminants eluting during the process based on the VP6 wild-type protein purification profile. The solubilised protein as per 3.7 and 3.8 sections, was injected into a pre-equilibrated HisTrap column with Equilibration buffer C. The column was washed with 10-column volumes of Equilibration buffer C to remove unbound protein. Elution of the targeted protein was eluted using a linear imidazole gradient, starting from 80 mM (Equilibration buffer C) to 600 mM (Elution buffer B) over 40-column volumes. This was followed by a final wash with 10-column volumes of 100% buffer B (Elution buffer B). The 2 ml fractions were collected throughout the elution gradient from F1 to F40 for further analysis.

Trial 2: Purification focused on removing contaminants during the elution process. Solubilized protein was injected into a pre-equilibrated HisTrap column with Equilibration buffer C. To remove weakly bound contaminants, the column was first washed with 10-column volumes of Wash buffer D containing 0.5% Tween-20, NaCl 150 mM. Using 0.5% Tween-20 as a wash step during protein purification is a reliable method for removing contaminants and non-specifically bound proteins while

maintaining the integrity of the target protein. This mild, non-ionic detergent efficiently disrupts hydrophobic interactions and releases loosely associated impurities without causing protein denaturation, making it a gentle yet effective choice for purification wash methods (Bornhorst and Falke, 2000). After the detergent wash, the column was washed with 20-column volumes of Equilibration buffer C to remove Tween-20 traces. The target protein was eluted using a linear imidazole gradient increasing from 80 mM (Equilibration buffer C) to 600 mM (Elution buffer B) over 40-column volumes, followed by an additional wash step using 10-column volumes of 100% Buffer B (Elution buffer B). The fractions (2 ml each) were collected throughout the gradient elution from F18 to F50 for further analysis.

Trial 3: Purification focused on removing contaminants during the elution process. Solubilized protein was injected into a pre-equilibrated HisTrap column with Equilibration buffer C. To further eliminate impurities, the column was washed with 10-column volumes of Wash buffer E (0.5% Tween-20, NaCl from 150 mM to 300 mM). This step effectively targeted and removed loosely bound contaminants. Following this, 20-column volumes of Equilibration buffer C was used to wash off the Tween-20 traces. Targeted proteins were eluted using a linear imidazole gradient increasing from 80 mM (Equilibration buffer C) to 600 mM (Elution buffer B) over 40-column volumes. This was followed by a final wash using 10-column volumes of 100% Buffer B (Elution buffer B). The fractions (2 ml each) were collected throughout the elution gradient from F18 to F80 for further analysis.

3.9.2 Analysis of Protein Purity Post-Purification

The purified protein aliquots were collected starting from the protein application flow-through, protein washes, and protein elution. These samples were analysed using the SDS-PAGE as per section 3.10.1. Based on the SDS-PAGE analysis, selected fractions were pooled together and concentrated using a 50 ml Millipore Amicon® Stirred Ultrafiltration Cell (Model 8050, Bedford, USA), fitted with a 10 kDa molecular weight cut-off Ultracel® membrane (Thermo Fisher Scientific, Waltham, MA, USA; PLGC Disk 10K, 44.5 mm). The ultrafiltration setup was placed on ice and stirred continuously to prevent protein denaturation by maintaining protein stability, with nitrogen gas pressure at 80 psi to facilitate the concentration process. After concentration, the protein samples were analysed using SDS-PAGE to determine if the proteins contained any contamination/ impurities before further analysis.

3.10 Gel electrophoresis of the VP6 proteins

3.10.1 SDS-PAGE Electrophoresis

SDS-PAGE electrophoresis was performed using the BioRad Mini-PROTEAN Electrophoresis Cell from Bio-Rad Laboratories, Inc (Hercules, California, United State), a system designed to separate proteins based on their molecular mass. The process uses three main components: 12% acrylamide/bis-acrylamide resolving gel [0.1 % (w/v) SDS, 0.05 % (w/v) APS, 0.1 % (w/v) TEMED as well as 0.38 M Tris-HCl buffer, pH 8.8], a 4% polyacrylamide stacking gel [0.1 % (w/v) SDS, 0.05 % (w/v) APS, 0.1 % (w/v) TEMED and 0.13 M Tris-HCl buffer, pH 6.8] and electrophoresis/running buffer (0.3% (w/v) Tris-HCl, 1.88% (w/v) glycine, and 0.1% (w/v) SDS pH 8.3). Protein separation was achieved through electrophoresis, based on the principle outlined by Laemmli, (1970), allowing molecular weight determination by comparing the migration distance of protein bands to known standards. An unstained protein ladder used was from Thermo Fisher Scientific, Cat. No. 26610 (Massachusetts, USA) consisting of seven reference protein sizes ranging from 14 kDa to 116 kDa: β-galactosidase (116 kDa), bovine serum albumin (66.2 kDa), ovalbumin (45 kDa), lactate dehydrogenase (35 kDa), restriction endonuclease Bsp98I (25 kDa), β-lactoglobulin (18.4 kDa) and lysozyme (14.4 kDa). These proteins were used for size estimation.

Protein samples were mixed with an equal volume of sample buffer containing 62.5 mM Tris-HCl pH 6.8, 10 % (v/v) glycerol, 10 % (v/v) SDS, 5 % (v/v) β-mercaptoethanol, and 0.05 % (w/v) bromophenol blue. Before loading, the mixtures were preheated at 95°C for 5 minutes to denature the proteins. A total volume of 10µl of each sample was loaded onto the SDS-PAGE gel alongside 5µl of the molecular weight marker. Electrophoresis was conducted at 140 V for 1 hour and 20 minutes. After the separation, gels were stained using Abcam InstantBlue Coomassie Protein Stain (Abcam, United Kingdom) and visualised using the BioRad Gel Dock Universal Hood (Hercules, United States of America). The protein band sizes were analysed by measuring the migration distance of both the protein of interest and the molecular weight marker using a ruler in centimetres (cm). The molecular weight standard curves were constructed in Microsoft Excel 2006 by plotting the migration distances of the proteins, and a linear regression trendline was used to calculate the approximate molecular weight of the VP6 protein band.

3.10.2 VP6 Quaternary Structure

3.10.2.1 Computational Prediction of the VP6 protein

AlphaFold 3 (an artificial intelligence (AI)-based software), was developed by DeepMind as a major advancement in the field of computational biology. Unlike AlphaFold 2, which primarily focuses on predicting single protein structures or protein–protein complexes, AlphaFold 3 models the entire biomolecular assemblies, including proteins, DNA/RNA, ligands, metal ions, and post-translational modifications. It accurately predicts complex 3D structures like protein–DNA or drug–protein interactions in a single model, making it a powerful tool for structural biology, drug discovery, and molecular design (Abramson *et al.* 2024).

The VP6 (consensus or wild-type) amino acid sequence obtained from Jailview database was formatted as three separate chains to model the trimeric assembly, and the same procedure was applied to the VP6 mutant sequence. The sequences were submitted to AlphaFold 3 (https://alphafoldserver.com/) accessed May 04, 2025 (Abramson *et al*. 2024) in multimer mode, with default settings for multiple sequence alignment (MSA) generation using MMseqs2. The model used AlphaFold 3's diffusion-based architecture, which directly predicts atomic coordinates for protein complexes. Multiple structural predictions were generated (five structural models per run), and the best model was selected based on confidence metrics including pLDDT (per-residue accuracy from 0-100 with plDDT > 90 as very high and plDDT < 50 as very low) and ipTM (interface predicted TM-score with values higher than 0.8 as high-quality predictions and values below 0.6 suggesting likely a failed prediction) (Zhang and Skolnick, 2004; Xu and Zhang, 2010). The final predicted model revealed a symmetric trimer with high-confidence interfacial contacts, consistent with known features of VP6 protein. In addition to the trimeric assembly, structural predictions were also performed for dimeric (2-chain), tetrameric (4-chain and alternative 4-chain configuration), pentameric (5-chain) and hexameric (6-chain) forms to assess potential oligomeric states.

3.10.2.2 Native-PAGE

Native PAGE was conducted using the BioRad Mini-PROTEAN Electrophoresis Cell from Bio-Rad Laboratories, Inc (Hercules, California, United State). This allowed for the separation of the VP6 wild-type protein and VP6 mutant protein based on their molecular mass and structural changes by involving the use of three main

components: 12% acrylamide/bis-acrylamide resolving gel [0.05 % (w/v) APS, 0.1 % (w/v) TEMED as well as 0.38 M Tris-HCl buffer, pH 8.8], the 5% polyacrylamide stacking gel [0.05 % (w/v) APS, 0.1 % (w/v) TEMED as well as 0.13 M Tris-HCl buffer, pH 7.5] and Bio-Rad's Tris/Glycine Native Running Buffer [0.303% (w/v) Tris-HCl, 1.441% (w/v) glycine, pH 8.3]. Before loading the protein, samples were combined with an equal volume of non-denaturant sample buffer (40% glycerol). A total of 20 µl per sample was loaded into the gel wells. Electrophoresis was performed at a constant voltage of 120 V for approximately 4 hours. Following electrophoresis, the gels were stained using Abcam InstantBlue Coomassie Protein Stain (Abcam, United Kingdom) to enhance band visibility. Visualization of protein bands was carried out using the Bio-Rad Gel Doc Universal Hood imaging system (Hercules, United States of America).

3.11 Purified VP6 Proteins Buffer Exchange and Concentration Determination

Buffer exchange involves the removal of small unwanted molecules such as NaCl, Imidazole, urea, and other small impurities from a concentrated protein solution through a semipermeable membrane, the buffers are exchanged in preparation for the protein functional assay (Rubick and Kreusch, 2018). The VP6 wild-type protein and VP6 mutant protein were processed with Elution buffer B containing 7M urea and subsequently concentrated following purification as per 3.9.2 section. The resulting concentrated proteins were transferred to snakeskin dialysis tubing and HiTrap Desalting. Several experimental trials were performed to determine optimal conditions that would minimize protein aggregation or precipitation. Various pH conditions were tested, specifically those within one unit of the isoelectric points of the VP6 wild-type protein (pI 6.63) and the VP6 mutant protein (pI 6.68) obtained from the Expasy ProtParam tool (https://web.expasy.org/protparam/) accessed on November 05, 2024. This optimization was crucial for maintaining the integrity of the protein during the buffer exchange and prior to concentration determination.

3.11.1 Buffer Exchange (Dialysis)

The resulting concentrated proteins were transferred to snakeskin dialysis tubing and immersed in dialysis buffer under continuous stirring to facilitate equilibration. Several experimental trials were performed to determine optimal conditions that would minimize protein aggregation or precipitation.

Trial 1: Three buffer exchange conditions were implemented, with freshly prepared buffer used for each condition: 4 hours with 4 M urea, 4 hours with 1 M urea, and 16 hours with 0 M urea, using Dialysis buffer A (50 mM sodium phosphate dibasic, pH 7.5, 7 M, 4 M, 1 M, and 0 M urea, and 0.02 % (w/v) sodium azide). Each buffer was used at a volume 20 times that of the protein sample to ensure complete submersion. These trials were conducted at a pH of 7.5, consistent with the same pH maintained during purification.

Trial 2: Three buffer exchange conditions were carried out: 4 hours with 4 M urea, 4 hours with 1 M urea, and 16 hours with 0 M urea, using Dialysis buffer B (50 mM sodium phosphate dibasic, pH 7.5, 20 mM NaCl, 7 M, 4 M, 1 M, 0 M urea, and 0.02 % (w/v) sodium azide) supplemented with 20 mM NaCl. Each buffer was used at a volume 20 times that of the protein sample to ensure complete submersion. These trials were performed at a pH of 7.5, consistent with the pH applied during the purification process.

Trial 3: Three buffer exchange conditions were implemented: 4 hours with 4 M urea, 4 hours with 1 M urea, and 16 hours with 0 M urea, using Dialysis buffer C (50 mM sodium phosphate dibasic, pH 5.5 and pH 8.0, 20 mM NaCl, 7 M, 4 M, 1 M, and 0 M urea, and 0.02 % (w/v) sodium azide). Each buffer was used at a volume 20 times that of the protein sample to ensure complete submersion. These trials were conducted at pH 5.5 and pH 8.0, representing acidic and basic conditions, respectively, to evaluate potential changes or improvements in protein environmental conditions.

Trial 4: Three buffer exchange conditions were implemented: 4 hours with 4 M urea, 4 hours with 2 M urea, and 16 hours with 0 M urea, using Dialysis buffer D (50 mM sodium phosphate dibasic, pH 7.5 containing 20 mM sodium chloride, 10% glycerol, 7 M, 4 M, 2 M, and 0 M urea, and 0.02 % (w/v) sodium azide). Each buffer was used at a volume 20 times that of the protein sample to ensure complete submersion. These trials were conducted at a pH of 7.5, consistent with the pH maintained during the purification process.

Trial 5: Three buffer exchange conditions were implemented under the same conditions: 4 hours with 4 M urea, 4 hours with 4 M urea, and 16 hours with 4 M urea, using Dialysis buffer E (50 mM sodium phosphate dibasic, pH 7.8 containing 20 mM sodium chloride, 7 M, 4 M, 4 M urea, and 0.02 % (w/v) sodium azide). Each buffer was

used at a volume 20 times that of the protein sample to ensure complete submersion. These trials were conducted at a pH of 7.8, chosen to be slightly alkaline, which may help to maintain protein solubility and stability while preventing aggregation, as proteins often tend to aggregate at lower pH values near their isoelectric points.

During dialysis trials, trial 1 through to trial 4 protein exhibited visible aggregation below 4 M urea concentration, marked by the formation of visible precipitates. At 0 M, a noticeable white precipitate was observed settling at the bottom of the snakeskin dialysis tubing. In contrast, Trial 5 at 4 M urea concentration-maintained protein integrity with no visible aggregation throughout the dialysis process.

3.11.2 Buffer Exchange (Desalting)

The concentrated VP6 wild-type protein and VP6 mutant protein samples were loaded into the Cytiva ATKA Super-loop, and they were injected into a HiTrap Desalting ready-to-use 5 mL column prepacked with Sephadex G-25. The column was pre-equilibrated with 50 ml of Dialysis buffer F (50 mM sodium phosphate dibasic, pH 7.8, containing 20 mM sodium chloride, 0 M urea, and 0.02 % (w/v) sodium azide). for efficient buffer exchange to the column. Here, the samples were allowed to pass through the Sephadex G-25 resin, where small molecules such as salts, imidazole, and urea were separated from the larger VP6 protein molecules. Buffer exchange occurred as the protein passed through the column, replacing the original buffer (Elution buffer B) with Dialysis buffer E. The elution was carried out using the same buffer at a flow rate of 1.5 ml/min. Fractions of 1 ml each were collected throughout the elution phase. The protein-containing fractions were pooled together after being identified by UV absorbance at 280 nm. These fractions containing the VP6 protein successfully desalted were used for further structural characterization same time to prevent aggregation as the proteins were observed to aggregate rapidly upon reaching low urea concentration, as confirmed in section 3.11.1.

3.11.3 Concentration Determination

The absorbance of the VP6 wild-type protein and VP6 mutant protein was measured using the Applied Photophysics Chirascan Plus, DichroWeb (Leatherhead, United Kingdom). Measurements were conducted using a 1 ml quartz cuvette. The VP6 protein concentrations were determined through spectrophotometric analysis using the EzDrop 1000c Micro-Volume Spectrophotometer (Blue-Ray Biotech, Taiwan).

After the concentration and dialysis of the VP6 wild-type protein and VP6 mutant protein, the concentrations were calculated by applying the Beer-Lambert Law, which correlates absorbance with concentration (Dadi and Yasir, 2022).

This method allows for exact measurement based on light absorption at specific wavelengths. The Beer-Lambert Law equation:

$$A = c.\varepsilon.l \quad (1)$$

In this equation, A represents the absorbance measured at 280nm, ε represents the molar extinction coefficient (M^{-1}. cm^{-1}), c is the protein concentration (M), and l represents the cuvette path length (which is 1 cm). When the extinction coefficient of a protein is unknown, it can be estimated or calculated using the method proposed by Pace *et al*. (1995), as shown in the equation below:

$$\varepsilon 280\ M\text{-}1.\ cm\text{-}1 = (5500 \times \Sigma W) + (1490 \times \Sigma Y) + (125 \times \Sigma C) \quad (2)$$

The extinction coefficients for both VP6 protein variants were calculated using the ExPASy ProtParam tool (https://web.expasy.org/protparam/) accessed on November 20, 2024, based on their amino acid sequences as described in Section 4.1 (Figures 4.1A and Figure 4.2A). The Tryptophan's (W) were 5, the Tyrosine's (Y) were 10 and the Cystines (C) were 3 in the VP6 protein sequences. The extinction coefficient for the VP6 wild-type protein and the VP6 mutant protein was 42400 M^{-1}. cm^{-1} if all cysteines were reduced and 42525 M^{-1}. cm^{-1} if all Cys residues form cystines. The selected extinction coefficient was 42400 M^{-1}. cm^{-1} for the VP6 protein with no cystines reduced. Observing the VP6 wild-type protein and VP6 mutation protein Sequences and structural data, the cysteines were not positioned to form disulfide bonds, suggesting that the cysteines do not participate in covalent cross-linking, which is consistent with the lack of disulfide bonds in VP6 protein.

The concentration of the VP6 proteins were confirmed using the Bradford assay and the bovine serum albumin (BSA) as the standard. BSA stock solution (2 mg/mL) was prepared using filtered Dialysis buffer E. Serial dilutions of this stock (ranging from 0.0625 to 0.2 mg/mL) were made using the same buffer to generate standard solutions. For the assay, 10 µl of either the VP6 protein samples or BSA standards were combined with 500 µl of the Bradford reagent (BioRad, Hercules, USA) and 490

µl of the nuclease-free H_20, resulting in 1 ml per sample. The mixtures were incubated for 10 minutes at room temperature before absorbance was measured using the EzDrop 1000c Micro-Volume Spectrophotometer (Blue-Ray Biotech, Taiwan). The absorbance at 595 nm was measured using a 1ml plastic cuvette for all prepared samples, and Microsoft Excel 2016 was used to plot the standard against the concentration. The trendline equation was used to calculate both protein concentrations as per section 4.22.

Following the measurement of absorbance and concentration for both the VP6 wild-type and mutant proteins, their respective yields were subsequently calculated.

$$Mass_{mg} = C_{\mu M} \times V_{ml} \times MW_{kDa} \quad (3)$$

Where M is mass, C for concentration, V for volume and MW for molecular weight.

3.12 Spectroscopy

3.12.1 Mass Spectrometry

3.12.1.1 In-gel Digestion

After protein purification as described in section 3.9.1, gel fragments containing the target proteins were cut, stored in an ultra-pure water and sent to CSIR (Pretoria, South Africa) for LC-MS analysis. Protein digestion was performed within gel fragments based on the methodology outlined by Shevchenko *et al.* (2006). Each gel fragment further divided into six smaller pieces. Proteins were reduced with 10 mM dithiothreitol (DTT) in 25 mM ammonium bicarbonate (NH_4HCO_3) at 60°C for one hour. This was followed by 10 minutes of incubation in 100% acetonitrile. Alkylation was then performed using Iodoacetamide (55 mM) in the dark. Then gel pieces were dehydrated and digested overnight 37°C using trypsin and chymotrypsin. To stop the digestion process, 0.1% formic acid was added. The resulting peptide mixtures were dried, then reconstituted in a solution of 2% acetonitrile and 0.2% formic acid prior to LC-MS analysis.

3.12.1.2 LC-MS Analysis

Peptides generated by trypsin and chymotrypsin digestion from each gel fraction were analysed using an Evosep One LC system coupled to a Bruker timsTOF HT mass spectrometer. Peptides were gradient-eluted and separated on a PepSep C18 column (150 µm × 8 cm, 1.5 µm particle size) with an acetonitrile gradient over 11.5 minutes.

The mass spectrometer operated in DDA-PASEF mode, scanning precursor ions within a mass range of 100-1700 m/z and acquiring data through 4 PASEF MS/MS ramps. MS/MS acquisition was triggered above an intensity threshold of 2500, with each scan accumulating data over approximately 100 milliseconds, resulting in a total cycle time of 0.53 seconds.

3.12.1.3 Data Analysis

Raw data files for VP6 wild-type and VP6 mutant protein were analysed using MaxQuant software with the Andromeda search algorithm. The search used an *Escherichia coli (E. coli)* sequence database from UniProt (Swiss-Prot and TrEMBL, **August 20, 2021**) and common contaminants. Trypsin and chymotrypsin were specified as enzymes for protein digestion, with cysteine alkylation (iodoacetamide) set as a fixed modification. Oxidation of methionine and N-terminal acetylation were included as variable modifications within the search parameters.

3.12.2 Characterization of the Secondary Structure

3.12.2.1 Circular dichroism

Far-UV circular dichroism (CD) spectroscopy was used to characterize both the VP6 wild-type protein and VP6 mutant proteins to analyse their secondary structure, folding, and binding characteristics.

Spectral measurements were conducted between wavelength of 190 nm and 250 nm using an Applied Photophysics Chirascan Plus, DichroWeb (Leatherhead, United Kingdom) in a 1 mm quartz cuvette. The instrument automatically measures the high voltage (HV) alongside Far-UV circular dichroism spectrum to confirm the reliability and quality of the data. The HV helps identify spectral regions with potential signal distortion due to high absorbance, where HV values higher than 600 V (signal-to-noise ratio) are considered poor and unreliable (Siligardi *et al.* 2018; Miles and Wallace, 2016). Prior to measuring any Far-UV circular dichroism spectrum, background correction was measured using Dialysis buffer E containing 50 mM sodium phosphate dibasic, 20 mM sodium chloride and 4 M, 5 M, 7 M urea concentration and Dialysis buffer F containing 50 mM sodium phosphate dibasic, 20 mM sodium chloride and 0 M urea, both adjusted to pH 7.8. The proteins were measured with a bandwidth of 1.5 nm every 1.0 nm scan interval in millidegrees, and the measurements were performed

in triplicate at 22°C. The obtained spectral data were converted to mean residue ellipticity using the following equation:

$$[\theta] = 100 \times (\text{signal})/c \times n \times l \quad (4)$$

In this equation, c represents the protein concentration in millimolar (mM), n represents the total number of amino acid residues in the protein, and l represents the pathlength of the cuvette in centimetres (cm) through which the light travels. This conversion standardizes the data, allowing for easier comparisons between proteins and the experimental conditions.

3.12.2.2 Attenuated Total Reflection (ATR)–Fourier Transform Infrared (FTIR) Spectroscopy

Attenuated total reflection Fourier transform infrared (ATR–FTIR) spectroscopy is a widely used method for analysing molecular composition and conformation of the structure. This technique is valuable for analysing protein secondary structures by detecting specific vibrational modes associated with amide bonds under various conditions (Kong and Yu, 2007; Krimm and Bandekar, 1986). The ATR-FTIR offers a unique advantage in studying structural changes in proteins, enabling real-time monitoring of secondary structure transitions without the need for sample labelling or modification and the technique is not time consuming.

In this study, the secondary structure of both VP6 wild-type protein and VP6 mutant proteins was characterized using a Nicolet™ iS^{TM}5 FTIR Spectrometer (Thermofischer Scientific) equipped with an iD7 ATR accessory, a ZnSe diamond crystal and a deuterated triglycine sulfate (DTSG) KBr detector. A 5 µl aliquot of each protein sample was analysed and isopropanol alcohol was used to clean the ATR crystal between measurements. The absorbance spectra were acquired using the Omnic Paradigm software (2019-2024, package version 1.9 Hot Fix 2 (Build 1) Thermofischer scientific). Each spectrum was obtained from 36 scans per sample at a resolution of 4 cm^{-1} between 525 cm^{-1} to 4000 cm^{-1} with air as the background. Background subtraction was performed using the Omnic Paradigm software, and the resulting data was exported in TSV format for further analysis.

3.12.3 Intrinsic Tryptophan Fluorescence

Fluorescence spectroscopy is an analytical technique that measures the emission properties of a molecule or protein with tyrosine and tryptophan residues by absorbing light, becoming excited, and then emitting light of a longer wavelength. This approach provides essential information about the molecular properties and behaviour of tryptophans in response to excitation; this is where the emission wavelength will shift depending on whether the tryptophans are buried inside a hydrophobic core or exposed to solvent. The VP6 wild-type protein and VP6 mutant protein, in Dialysis buffer E containing 50 mM sodium phosphate dibasic and 4 M, 5 M and 7 M urea concentration and Dialysis buffer F containing 50 mM sodium phosphate dibasic and 0 M urea concentration at pH 7.8, were characterized by using intrinsic fluorescence, which focuses on the tertiary structure, specifically targeting tryptophan (Trp) residues. Tryptophan fluorescence often dominates in proteins having both aromatic residues, and it is more sensitive to microenvironmental changes than tyrosine because proteins frequently have few tryptophan residues. Fluorescence experiments primarily detect these sites (Ghisaidoobe and Chung, 2014). The fluorescence measurements were conducted using an Applied Photophysics Chirascan Plus (Leatherhead, United Kingdom) with a 1 mm quartz cuvette. Emission spectra were measured between wavelengths of 280 nm to 500 nm using a bandwidth of 1.5 nm. Tryptophan's were selectively excited at 295 nm, a wavelength chosen to minimize the excitation of other aromatic residues such as tyrosine and phenylalanine. Emission was observed between the wavelengths of 335 nm to 350 nm. All spectra were collected in triplicate.

3.12.4 Thermal Unfolding

Thermal unfolding of both VP6 wild-type protein and VP6 mutant protein was monitored using far-UV circular dichroism and tryptophan fluorescence spectroscopy in Dialysis buffer E (50 mM sodium phosphate dibasic and 4 M urea at pH 7.8). For both far-UV circular dichroism and tryptophan fluorescence spectroscopy, the temperature was gradually increased from 20 °C to 94 °C, cooled back to 20 °C using the temperature controller instrument PCS.3 Single Cell Peltier Temperature Controller, with a circulating chiller unit programmed by the Chirascan's Pro-Data software (Leatherhead, United Kingdom). An additional measurement was conducted after the same sample was stored overnight at 4 °C. The procedures for far-UV circular dichroism and tryptophan fluorescence spectroscopy measurements followed the

protocol outlined in sections 3.12.2.1 and 3.12.3, respectively. Measurements for both far-UV circular dichroism (CD) and tryptophan fluorescence spectroscopy were recorded at 1 °C intervals during the temperature ramp to monitor structural changes. These measurements were designed to determine whether the thermal denaturation of the VP6 protein undergoes complete or irreversible unfolding and whether any refolding occurred upon cooling to 20 °C and overnight at 4 °C after the thermal ramp.

4 Results

4.1 Protein Consensus Sequence Identification and Engineering

The sequence alignment of 2809 Rotavirus A (RVA) VP6 sequences yielded a consensus sequence that represented the most common amino acid residues found at each location of this protein. Through this analysis, 30 amino acid regions were identified that displayed variability across the consensus sequences (Figure 4.1A highlighted in red ●). The 30 amino acids had limited amino acid substitution(s) at these sites, revealing how different RVA strains have slight changes in their VP6 sequences (Figure 4.1B). The consensus sequence was analysed using the AGGRESCAN3D tool to determine the aggregation prone regions, and this was correlated with the 30 amino acid regions displaying variability. Six amino acids (Figure 4.2A highlighted in yellow ● position 39, 109, 151, 248, 252, and 281) displayed a high probability of aggregation and could be mutated to amino acids that occurred naturally at these positions. Specific amino acid substitutions were chosen to improve the aggregation score (Figure 4.2B). During this process, Proline (P) was avoided as a substitution due to it being structurally rigid, which can disrupt the formation of secondary structures like alpha helices and beta sheets by either hindering normal folding, resulting in protein aggregation, misfolding, or abnormal function (Imai and Mitaku, 2005). The final predicted sequence (Figure 4.2A highlighted in yellow ●) was obtained by altering those six amino acids marked in yellow, which AGGRESCAN3D identified as "hot spots" for aggregation.

The selection of each amino acid was also incorporated with the AlphaFold tool, which was used to predict the structure representing the consensus sequence (Figure 4.3A). Following the prediction, the results were obtained in a zip file containing multiple models, predictions, ranking values, and corresponding PDB files. From the five models generated, model number 4 was selected based on its rank number 001, demonstrating the closest resemblance to the 1QHD structure according to the local distance difference test Cα (lDDT-Cα) estimated by the pLDDT (Figure 4.3B). The test compared positions of Cα atoms between the predicted models and the known structure, with lower differences indicating higher accuracy.

A

>Consensus/1-397 Percentage Identity Consensus

MEVLYSLSKTLKDARDKIVEGTLYSNVSDLIQQFNQMIVTMNGNDFQTGGIGNLPIRNWTFDFGLLGT
TLLNLDANYVETARTTIEYFIDFIDNVCMDEMARESQRNGVAPQSEALRKLAGIKFKRINFNNSSEYI
ENWNLQNRRQRTGFVFHKPNIFPYSASFTLNRSQPMHDNLMGTMWLNAGSEIQVAGFDYSCALNAPAN
IQQFEHIVQLRRALTTATITLLPDAERFSFPRVINSADGATTWFFNPIILRPNNVEVEFLLNGQIINT
YQARFGTIVARNFDTIRLSFQLMRPPNMTPAVNALFPQAQPFQHHATVGLTLRIESAVCESVLADANE
TLLANVTAVRQEYAIPVGPVFPPGMNWTELITNYSPSREDNLQRVFTVASIRSMLIK

B

Positions	Amino acid in consensus	Amino acid residues located at this position	Positions	Amino acid in consensus	Amino acid residues located at this position
2	E	D	172	M	Q, A, and P
39	V	I, L, and A	199	L	I, F, and T
45	D	E, N, K, and G	205	I	T and V
60	T	N, A, and S	217	A	V, P, and L
83	T	N	248	F	Y, H, and L
86	E	D	252	I	V, A, and L
89	I	V	281	V	I, T, M and L
92	I	V	305	N	A, K, and T
101	A	V	310	Q	N
109	V	I and P	315	Q	E and I
114	E	G	339	N	S and G
115	A	S	342	L	M and T
120	A	S	348	A	S, Y, and G
130	N	D and K	369	E	D
151	V	T, P, A, and I	396	I	V

Figure 4.1: VP6 Consensus Sequence Optimisation

Shows the predicted consensus sequence of the VP6 protein in FASTA format **(A)**, generated using Jailview software revealing 30 amino acid regions with heightened variability, highlighted in red and one-letter codes for amino acids were used. **(B)** The table shows increased variability across specific positions and the amino acid substitution(s) at these sites, revealing changes in different RVA VP6 strains.

A

>Consensus/1-397 Aggrescan Mutated Final Sequence

MEVLYSLSKTLKDARDKIVEGTLYSNVSDLIQQFNQMIATMNGNDFQTGGIGNLPIRNWTFDFGLLGT
TLLNLDANYVETARTTIEYFIDFIDNVCMDEMARESQRNGLAPQSEALRKLAGIKFKRINFNNSSEYI
ENWNLQNRRQRTGFTFHKPNIFPYSASFTLNRSQPMHDNLMGTMWLNAGSEIQVAGFDYSCALNAPAN
IQQFEHIVQLRRALTTATITLLPDAERFSFPRVINSADGATTWHFNPTILRPNNVEVEFLLNGQIINT
YQARFGTITARNFDTIRLSFQLMRPPNMTPAVNALFPQAQPFQHHATVGLTLRIESAVCESVLADANE
TLLANVTAVRQEYAIPVGPVFPPGMNWTELITNYSPSREDNLQRVFTVASIRSMLIK

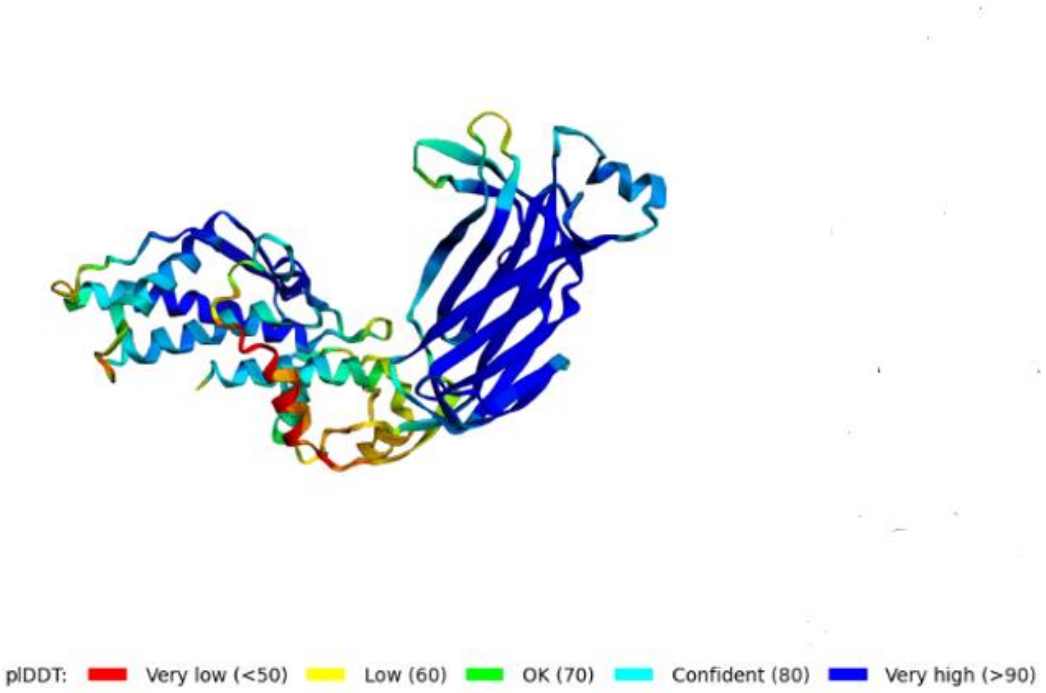
B

Amino acids	Aggrescan3D score before	Agrescan3D score after		Amino acids	Aggrescan3D score before	Agrescan3D score after
Amino acid (V) Position A39				**Amino acid (F) Position A248**		
V-I	1.3439	1.4868		F-Y	1.9053	1.1786
V-L	1.4868	1.0411		F-H	1.9053	-1.2161
V-A	1.0589	-0.4562		F-L	1.9053	1.0765
Amino acid (V) Position A109				**Amino acid (I) Position A252**		
V-I	0.4296	0.0992		I-V	1.7073	0.9212
V-L	0.4296	-0.3505		I-A	1.7073	0.9212
V-P	0.4296	-1.2069		I-T	1.0732	0.5742
Amino acid (V) Position A151				**Amino acid (V) Position A281**		
V-T	1.1153	-0.6973		V-I	0.7704	0.6621
V-P	1.1153	-0.8959		V-T	0.7704	-0.6117
V-A	1.1153	-0.5301		V-M	0.7704	0.004
V-I	1.1153	1.3566		V-A	0.7704	-0.5052

Figure 4.2: VP6 Mutated Sequence and Aggrescan3D Score

The final predicted VP6 sequence was optimised using AGGRESCAN3D tool by incorporates six amino acid substitutions (highlighted in yellow) that significantly reduce aggregation propensity **(A)**. These substitutions were selected based on reduced aggregation scores, enhancing solubility and structural stability. **(B)** The table lists the aggregation scores of the six amino acids identified by Aggrescan3D as being prone to aggregation. Residues highlighted in yellow represented the selected substitutions that significantly reduce aggregation propensity and improve protein solubility.

A



B

```
2024-02-15 20:02:33,716 reranking models by 'plddt' metric
2024-02-15 20:02:33,716 rank_001_alphafold2_ptm_model_4_seed_000 pLDDT=80.8 pTM=0.571
2024-02-15 20:02:33,718 rank_002_alphafold2_ptm_model_1_seed_000 pLDDT=80.2 pTM=0.56
2024-02-15 20:02:33,719 rank_003_alphafold2_ptm_model_2_seed_000 pLDDT=79.5 pTM=0.57
2024-02-15 20:02:33,719 rank_004_alphafold2_ptm_model_3_seed_000 pLDDT=79.4 pTM=0.517
2024-02-15 20:02:33,720 rank_005_alphafold2_ptm_model_5_seed_000 pLDDT=78.4 pTM=0.508
```

Figure 4.3: VP6 Protein Structural Prediction

Shows the structure prediction using the consensus sequence to analyse the structural integrity and confidence levels of various regions within the predicted VP6 mutant protein structure form AGGRESCAN3D tool, using data from the AlphaFold database. **(A)** The three-dimensional structure was visualized to highlight regions based on the Local Distance Difference Test (LDDT) score, where regions with a score drop of 50% are portrayed with low confidence, while regions with a score of 90% are portrayed with high confidence. **(B)** The rankings of the five models predicted by the Aphafold database presented per-residue confidence metric known as the predicted LDDT (pLDDT). This selection process ensures the use of the most reliable model for further analyses and interpretations, facilitating a more accurate representation of the VP6 protein's predicted structure.

The pLDDT score supports this theory since it estimates the confidence of the model's structural accuracy, where a higher pLDDT score means a more reliable model (Jumper *et al.* 2021). Model 4 revealed the highest Local Distance Difference Test (LDDT) score of 80.8 (Figure 4.3B), determined based on the per-residue confidence metric scale where a score drop of 50% represented low confidence, and regions scoring 90% represented high confidence.

To ensure fidelity to the protein's function, amino acid substitutions were carefully selected, avoiding interference with protein structure and functionality. The final sequence was reverse translated into DNA, specifically the VP6 gene encoding sequence, using the NovoPro Bioscience Codon Optimization Tool (Expoptimizer) to facilitating downstream molecular biology applications. To assess the protein sequence stability index of the final predicted VP6 mutant sequence, the Expasy ProtParam tool was used, and its stability was compared with that of the VP6 wild-type sequence obtained from Jailview. The stability index of the VP6 consensus sequence and the final predicted sequence were determined by systematically testing the amino acid sequence before mutation and the final amino acid substitutes. The VP6 consensus sequence displayed an instability index of 42.30. Similarly, the final mutated sequence displayed an instability index of 43.59. These stability values showed that the 6 mutated amino acids did not alter the stability of the protein drastically and it was deemed acceptable, ensuring a balance between structural integrity, aggregation propensity and stability.

4.2 Bacterial Growth Curve and VP6 Protein Overexpression

4.2.1 Bacterial Growth Curve

When bacteria are introduced into a fresh medium in a closed system, their growth follows a characteristic pattern. During the lag phase, bacteria start to grow slowly, adjusting to the new environment, increasing in size without significant replication (Rolfe *et al.* 2012). In the log (exponential) phase, cells begin dividing through binary fission, reaching the maximum growth rate. As nutrients become limited, the culture enters the stationary phase, where cell division balances cell death.

Finally, in the death phase, nutrient depletion and waste accumulation lead to a decline in viable cells, with dead cells outnumbering living ones as bacteria lose the ability to divide (Madigan *et al.* 2015). The choice of OD_{600} 0.4 to 0.6 as the reference point for induction was crucial, as it signifies a stage in bacterial growth where cells are actively growing at the fastest rate and multiplying.

After successful transformation and selection, the growth of BL21(DE3) cell line with the plasmid encoding VP6 wild-type protein highlighted in orange ● and the VP6 mutant protein highlighted in green ● were monitored. The BL21 (DE3) bacterial cells required approximately 70 minutes to progress to the early-log growth phase (Figure 4.4) with OD_{600} readings of 0.669 highlighted in red ● and 0.560 highlighted in yellow ● respectively. This slight difference in optical density is minimal and likely due to experimental variability during OD_{600} measurements, rather than a biologically significant difference, making them ideal as induction points. Therefore, the data do not support a clear distinction in growth behaviour between the two constructs in BL21(DE3) cells.

Whereas Figure 4.5 shows the NiCo21 (DE3) cell line with plasmid encoding VP6 wild-type protein highlighted in orange ● and VP6 mutant protein highlighted in green ● showing different growth rates. Cells containing plasmid encoding VP6 wild-type protein, requiring approximately 80 minutes to reach early-log growth phase, with induction point at OD_{600} of 0.601 highlighted in red ●. Whereas cells containing plasmid encoding VP6 mutant protein required approximately 70 minutes to reach early-log growth phase with OD_{600} of 0.581 highlighted in yellow ● as the induction point. Even though there was a slight difference in timing and optical density, these changes were relatively small and may not be biologically meaningful. Given that the VP6 mutant protein differs from the VP6 wild-type protein by only six amino acids, it is unlikely that this alone accounts for the difference.

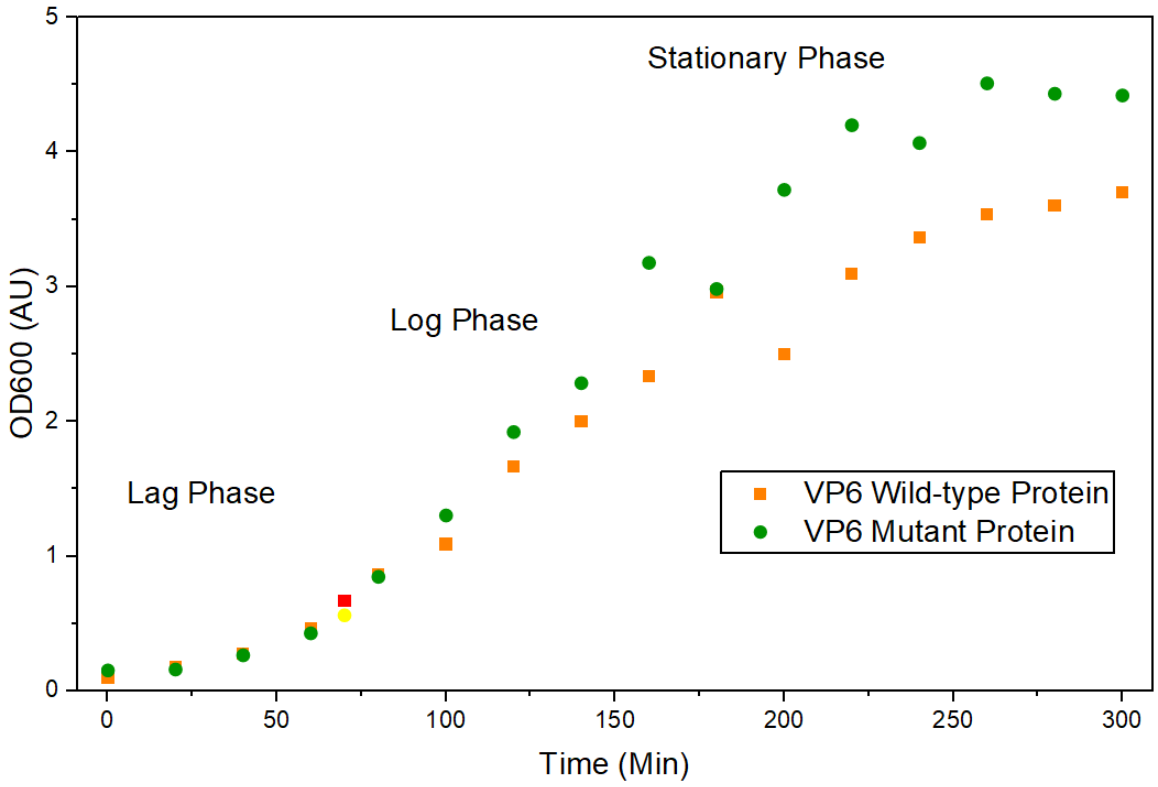


Figure 4.4: BL21(DE3) with VP6 Encoding Plasmid Growth Curve

The growth curve of BL21 (DE3) cell lines transformed with the plasmids (encoding VP6 wild-type protein highlighted in orange ● and VP6 mutant protein highlighted in green ●) were measured over time. The optimal IPTG induction points were identified at an OD_{600} of 0.669 (red) for the VP6 wild-type and 0.56 (yellow) for the VP6 mutant, indicating the optimal cell densities for induction.

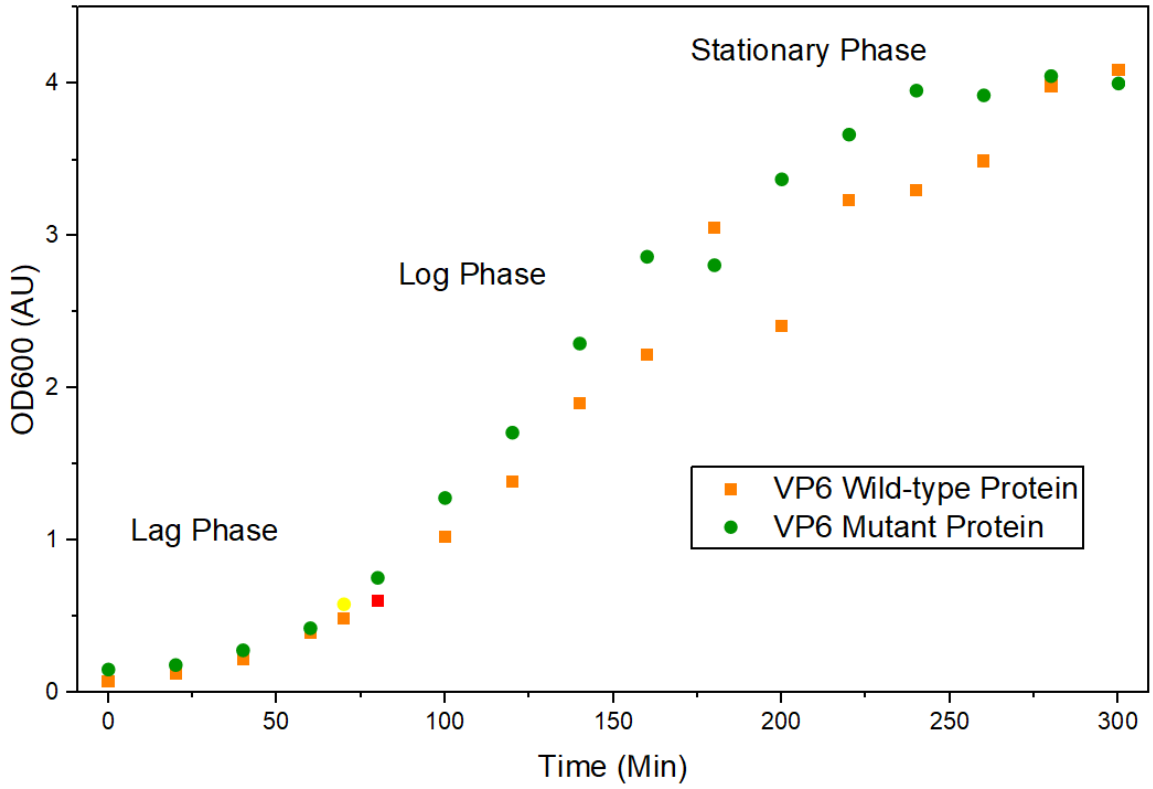


Figure 4.5: NiCo21 (DE3) with VP6 Encoding Plasmid Growth Curve

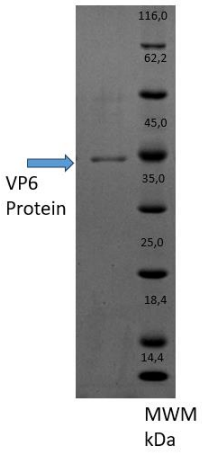
Shows the NiCo21(DE3) cell lines transformed with the plasmids (encoding VP6 wild-type protein highlighted in orange ● and VP6 mutant protein highlighted in green ●. were monitored over time. The optimal IPTG induction times were found at OD_{600} values of 0.601 (red) for the VP6 wild-type and 0.581 (yellow) for the VP6 mutant, showing the optimal cell densities for induction.

4.2.2 VP6 Protein Expression

Cell lines BL21(DE3) and NiCo21(DE3) were used to express the VP6 consensus (wild-type) encoding gene and VP6 mutant encoding gene cloned into pET15b, which carried an Ampicillin resistance gene. Their overexpression behaviours were compared between the two *E. coli* expression strains under various induction conditions. The goal was to identify a cell line and conditions yielding the highest amount of VP6 protein, preferably soluble, minimizing the time and resources required. The conditions tested during the VP6 protein expression included temperature (20°C and 37°C), IPTG concentration (BL21 (DE3): 1.0 mM, 0.1 mM and NiCo21 (DE3): 0.4 mM, 0.04 mM), and induction time (0, 4, 5, 6, 7, 8, and 12 hours). After sonication, cell samples were analysed using SDS-PAGE.

The relative molecular weight of the VP6 wild-type protein was determined to be approximately 48.41 kDa as shown in Figure 4.6. This estimation was based on the migration distances of the protein bands on the SDS-PAGE gel relative to the molecular weight marker from the purification trials. The molecular weight of the VP6 mutant protein was calculated to be approximately 47.86 kDa, as shown in Figure 4.7. Figure 4.6A (VP6 wild-type protein) and Figure 4.7A (VP6 mutant protein) were included to illustrate the migration profile of the purified proteins. The migration distance for VP6 wild-type (Figure 4.6B) and VP6 mutant protein (Figure 4.7B) were measured using a ruler in cm for the standard curve, providing a visual marker for its estimated molecular weight. In these figures, the purified protein bands were clearly labelled, serving as visual confirmation of the VP6 proteins. The expressed, solubilised and purified VP6 proteins (VP6 wild-type protein and VP6 mutant protein) as observed from all the SDS-PAGE gels from sections 4.2 to section 4.5 were also measured and consistently displayed molecular weights that were approximately 3 kDa higher than the VP6 protein mass of 45 kDa which was experimentally confirmed and reported by Bugli *et al.* 2014. Additionally, Gasteiger *et al*. 2005 theoretically confirmed the VP6 protein mass as 45 kDa. The difference in molecular weight/mass might be due to the Histidine tag containing 6 histidine at the N-terminal, which adds additional mass to the recombinant proteins. Such tags are commonly used for affinity purification but may cause slight increases in apparent molecular weight when analysed by SDS-PAGE.

A



B

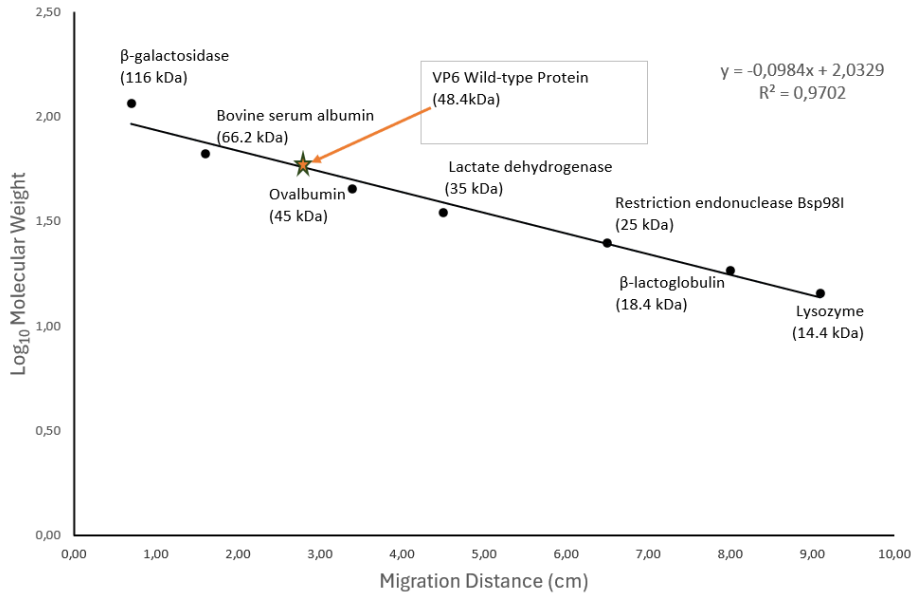
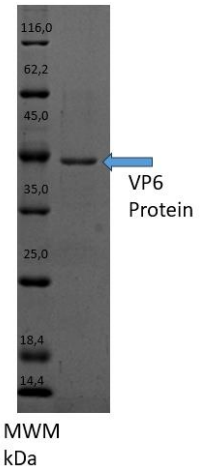


Figure 4.6: VP6 Wild-type Protein Molecular Weight Calculation

To estimate the molecular weight of the VP6 wild-type protein, a calibration curve was constructed by comparing its migration distance to that of the molecular weight marker. **(A)** The analysis was carried out using the overexpressed and purified protein, alongside MWM. **(B)** The migration distance was measured in centimetres (cm), and the corresponding molecular weights were calculated using a $\log_{10}$ equation. The calculated molecular weight of VP6 was approximately 48.4 kDa. The migration position of the VP6 protein was marked with an orange star on the standard curve, providing a visual marker for its estimated molecular weight.

A



B

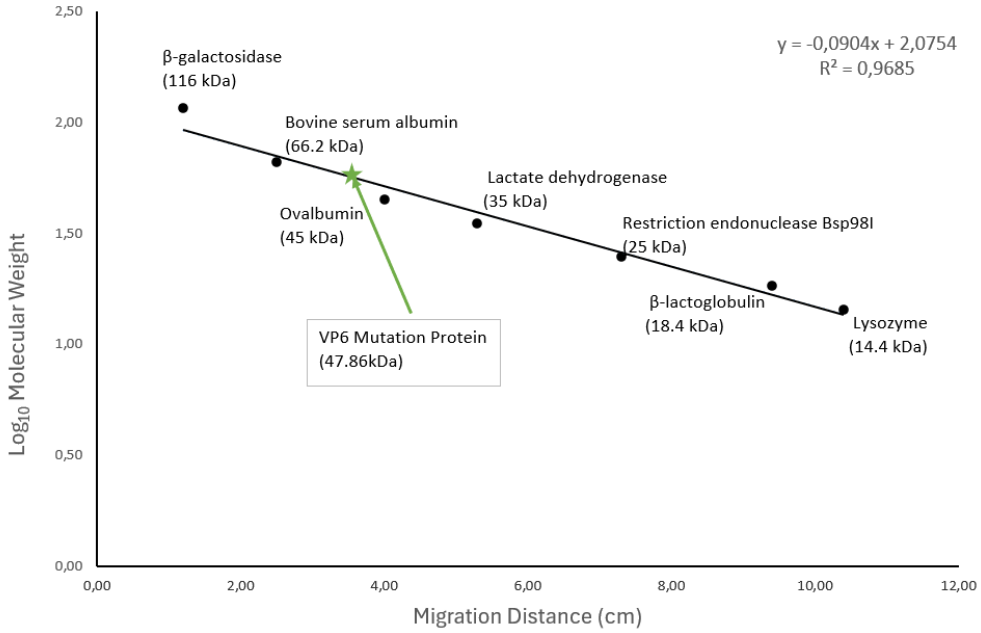


Figure 4.7: VP6 Mutant Protein Molecular Weight Estimation

A calibration curve was generated to determine the molecular weight of the VP6 mutant protein, a standard calibration curve was constructed by comparing its migration distance to that of the MWM. **(A)** The analysis was conducted using the overexpressed and purified protein alongside MWM. **(B)** The migration distances were measured in centimetre (cm), and the molecular weights were calculated using a $\log_{10}$ equation. The VP6 mutant protein molecular was calculated to be approximately 47.86 kDa. The migration position of VP6 protein was marked with a green star on the standard curve, serving as a visual reference point for its estimated size.

4.2.2.1 Overexpression of the VP6 Wild-type Protein

The VP6 wild-type protein was successfully overexpressed in *E. coli* strains BL21(DE3) and NiCo21(DE3) at both 20°C and 37°C. The expression occurred predominantly in the insoluble fraction, indicating the formation of inclusion bodies with a small amount of protein present in soluble form (Figure 4.8 and Figure 4.9).

Using the BL21(DE3) cell line, VP6 wild-type protein expression was observed at both 20°C and 37°C using IPTG concentrations of 0.1 mM and 1.0 mM. At 20°C, the overexpression was successful, mainly in insoluble form, with a small amount of soluble protein in the supernatant, increasing over time. Lower IPTG concentrations (0.1 mM) resulted in slower cellular growth at 20°C (Figure 4.8A), and the protein expression was similar even with higher IPTG induction (1.0 mM) (Figure 4.8B). This suggested that time rather than temperature or IPTG concentration had a greater influence on the protein yield with the highest expression observed after 12 hours. At 37°C, overexpression was successful with no substantial increase in protein yield after 7 hours post-induction, regardless of IPTG concentration of 0.1 mM (Figure 4.8C) and 1.0 mM (Figure 4.8D). An increase in temperature might have stressed the host cells, reducing overall viability and productivity. Different cellular growth was observed at both temperatures after induction, suggesting a possible burden or cell toxicity associated with the overexpression of the protein. This was reported by Mital *et al.* (2021), who confirmed that even though *E. coli* BL21(DE3) is the most used for recombinant protein overexpression, especially from strong T7 promoters, it can result in metabolic stress or toxicity, particularly in those recombinant proteins that form inclusion bodies.

With NiCo21(DE3), VP6 overexpression was observed at both 20°C and 37°C using different IPTG concentrations (0.04 mM and 0.4 mM). The protein was mainly in insoluble form with a slight increase in soluble protein over time. At 20°C, the expression was time-dependent for both 0.04 mM (Figure 4.9A) and 0.4 mM (Figure 4.9B) IPTG concentration. Low IPTG concentration produced higher protein yield after 12 hours post induction. While 12 hours was the maximum duration tested, it was unclear whether protein expression would continue beyond this point.

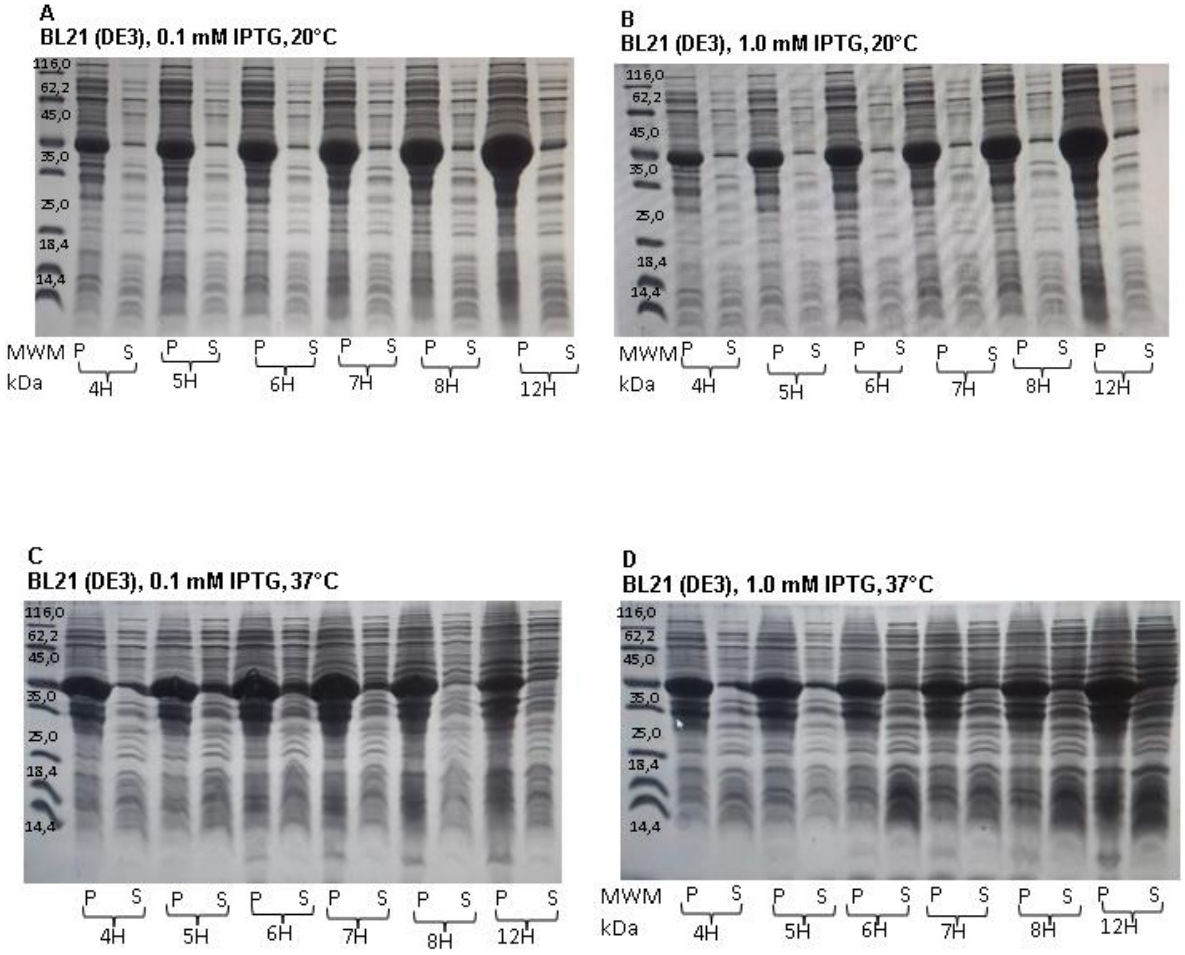


Figure 4.8: VP6 Wild-type Protein Expression using BL21 (DE3)

Shows the overexpressed of the VP6 wild-type protein using BL21 (DE3) at 20°C (**A** and **B**) and 37°C (**C** and **D**). Lanes were represented by P=pellet, S=supernatant and MWM= Molecular weight marker size in kDa. The BL21 (DE3) with the VP6 wild-type protein was induced with IPTG of 0.1mM and 1.0 mM. The cell line was grown from 4H, 5H, 6H, 7H, 8H, and 12 H.

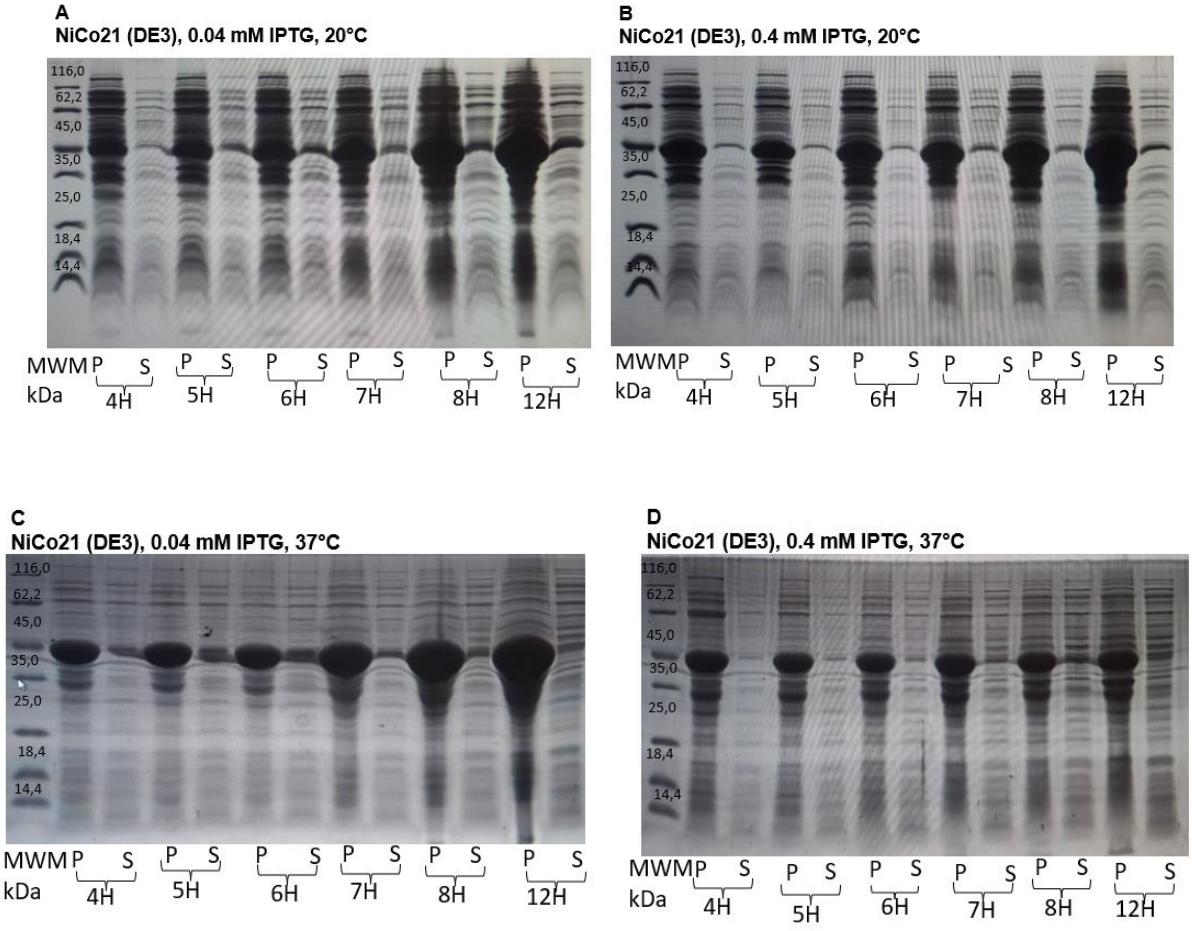


Figure 4.9: VP6 Wild-type Protein Expression using NiCo21 (DE3)

Shows the overexpressed of the VP6 wild-type protein using NiCo21 (DE3) at 20°C (**A** and **B**) and 37°C (**C** and **D**). Lanes were represented by P=pellet, S=supernatant and MWM= Molecular weight marker size in kDa. The BL21 (DE3) with the VP6 wild-type protein was induced with IPTG of 0.04mM and 0.4 mM. The cell line was grown from 4H, 5H, 6H, 7H, 8H, and 12 H.

At 37°C, protein expression was notably more rapid and reached higher yields at earlier time points, again accumulating predominantly in the insoluble fraction. Induction with 0.04 mM IPTG concentration produced the highest expression after 12 hours (Figure 4.9C), while 0.4 mM IPTG concentration resulted in relatively stable expression over time (Figure 4.9D). NiCo21(DE3), an engineered derivative of BL21(DE3), was used for its ability to enhance the purity of His-tagged proteins by reducing background binding during IMAC purification (Robichon *et al.* 2011).

4.2.2.2 Overexpression of the VP6 Mutant Protein

The overexpression of the VP6 mutant protein was successful in both *E. coli* BL21(DE3) and NiCo21(DE3) strains at 20°C and 37°C. However, in all conditions tested, the expressed protein was predominantly in the insoluble from (Figure 4.10 and Figure 4.11).

The VP6 mutant protein was successfully overexpressed in BL21 (DE3) cell line at 20°C and 37°C using IPTG concentrations of 0.1 mM and 1.0 mM. At 20°C, the overexpression levels were similar for both IPTG concentrations of 0.1 mM (Figure 4.10A) and 1.0 mM (Figure 4.10A), with protein yield increasing over time and reaching a maximum at 12 hours post-induction. This suggested that lower temperature slowed cellular metabolism and protein synthesis, allowing gradual accumulation of the protein regardless of IPTG concentrations. At 37°C, overexpression remained high across all conditions. Notably, 0.1 mM IPTG concentration induced slightly higher protein levels at earlier time points (5 to 7 hours) (Figure 4.10C), while 1.0 mM IPTG concentration showed relatively stable protein expression throughout the 12-hour induction period (Figure 4.10D). These observations suggest that lower IPTG concentration may reduce metabolic burden, supporting higher early yields, whereas higher IPTG may induce stress, limiting further increases over time.

With the NiCo21(DE3) strain, the protein expression was tested using IPTG concentrations of 0.04 mM and 0.4 mM at both 20°C and 37°C. At 20°C, protein expression level increased over time for both IPTG concentrations 0.04 mM (Figure 4.11A) and 0.4 mM (Figure 4.11B), reaching maximum yield at 12 hours post-induction. However, the low temperature did not significantly improve solubility, indicating potential folding issues inherent to the mutant protein, as reported by Francis and Page, (2010).

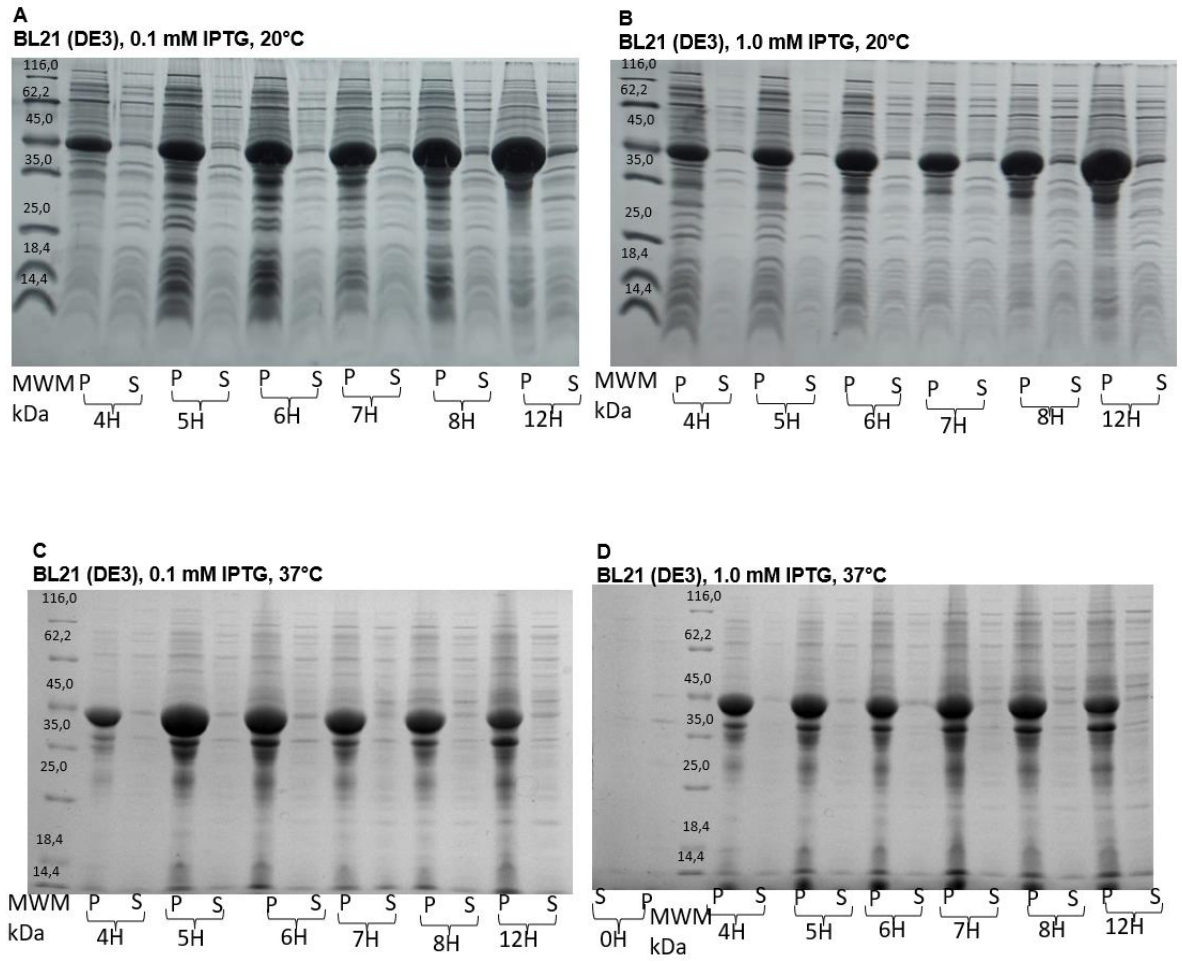


Figure 4.10: VP6 Mutant Protein Expression using BL21 (DE3)

Shows the overexpressed of the VP6 mutant protein using BL21 (DE3) at 20°C (**A** and **B**) and 37°C (**C**and **D**). Lanes were represented by P=pellet, S=supernatant and MWM= Molecular weight marker size in kDa. The BL21 (DE3) with the VP6 wild-type protein was induced with IPTG of 0.1mM and 1.0 mM. The cell line was grown from 0H, 4H, 5H, 6H, 7H, 8H, and 12 H.

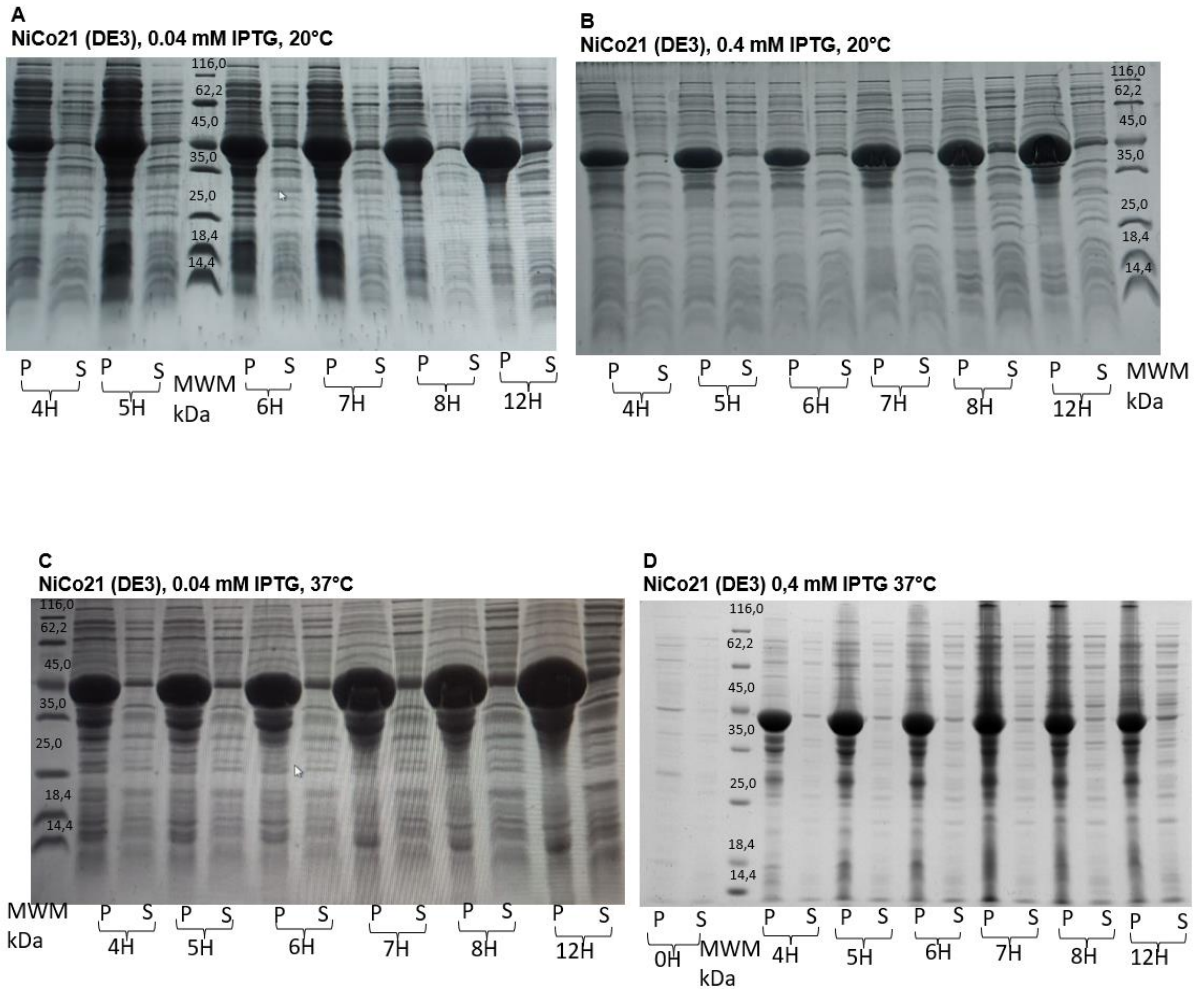


Figure 4.11: VP6 Mutant Protein Expression Using NiCo21 (DE3)

Shows the overexpressed of the VP6 mutant protein using NiCo21 (DE3) at 20°C (**A** and **B**) and 37°C (**C** and **D**). Lanes were represented by P=pellet, S=supernatant and MWM= Molecular weight marker size in kDa. The BL21 (DE3) with the VP6 wild-type protein was induced with IPTG of 0.04 mM and 0.4 mM. The cell line was grown from 0H, 4H, 5H, 6H, 7H, 8H, and 12 H.

At 37°C, similar time-dependent increases in expression were observed. With 0.04 mM IPTG concentration producing a high yield of protein from 7 hours onward, with the highest protein yield observed at 12 hours post-induction (Figure 4.11C). Lower IPTG concentrations likely reduced metabolic burden, enabling sustained expression. However, at 0.4 mM IPTG concentration, protein levels remained relatively stable throughout the induction period, possibly due to cellular stress or early saturation of the expression machinery (Figure 4.11D).

Both VP6 wild-type protein and VP6 mutant protein were successfully overexpressed in *E. coli* strains BL21(DE3) and NiCo21(DE3), primarily in insoluble form. Among tested conditions, NiCo21(DE3) at 37°C with 0.04 mM IPTG concentration consistently yielded the highest expression levels for both proteins at 12 hours post-induction. Lower IPTG concentrations reduced metabolic stress and proved cost-effective without compromising yield. These results highlight NiCo21(DE3) as a more suitable host for VP6 expression and provide a foundation for future efforts to improve solubility and refolding.

4.3 Isolation of the VP6 Protein Inclusion Bodies and Solubilisation

The formation of inclusion bodies is a significant challenge in the overexpression, solubilization, and purification of recombinant proteins when using *E. coli* as a host system (Burgess, 2009). The wash steps on both VP6 proteins assisted in the removal of the cell debris from the inclusion bodies. After the Triton X-100 and the Tris-HCl washes in both VP6 wild-type protein (Figure 4.12A) and VP6 mutant protein (Figure 4.12B), it was observed that a substantial amount of non-targeted proteins was reduced during the wash steps in the pellet, preparing the washed pellet for further solubilisation using different urea concentrations.

This additional step of inclusion bodies isolation was added as a step to improve the VP6 protein purity by successfully removing the contaminants, membrane fragments, and other cell debris that bind alongside the protein. Inclusion bodies often trap both the target protein and cellular debris or tightly bound contaminants. Washing these beforehand helps remove impurities that standard washes can't eliminate, improving the overall purity before chromatography even begins.

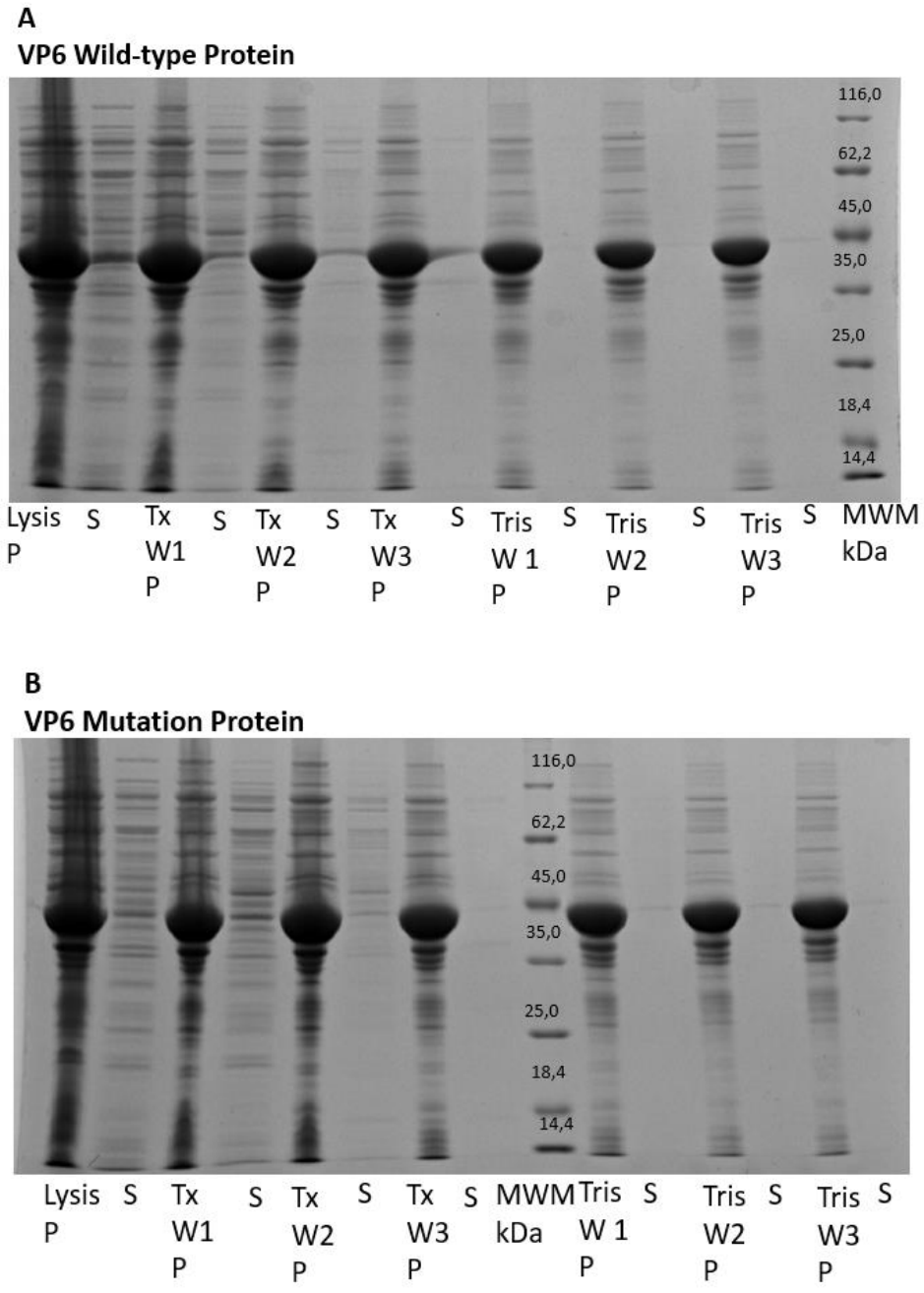


Figure 4.12: Isolation of Inclusion Bodies for VP6 Protein

Shows the successful isolation of the inclusion bodies from the overexpressed VP6 protein using the NiCo21 (DE3). The VP6 wild-type protein **(A)** and the VP6 mutant protein **(B)** shows successful washes of the inclusion bodies with Triton X-100 and Tris-HCl at pH 7.8. lanes were presented as Tx W = Triton x-100 wash, the Tris W=Tris-HCl wash, P=pellet, S=supernatant, and MWM= molecular weight marker.

Additional wash steps were included during purification as per section 4.5, to remove contaminants co-eluted with the target protein and ensure VP6 purity. Both VP6 wild-type protein and VP6 mutant protein showed similar outcomes, highlighting the importance of washing inclusion bodies to eliminate cell debris and membrane fragments. This optimized protocol improved the overall quality and yield of VP6 protein, facilitating more efficient downstream processing and analysis.

4.4 The VP6 Protein Solubilisation

To produce bioactive proteins, inclusion bodies must go through substantial processing, which includes solubilisation and purification. The solubilisation yield was assessed by determining the impact of urea concentration on protein solubility to improve protein recovery. For the VP6 wild-type protein (Figure 4.13A) and VP6 mutant protein (Figure 4.13B), Solubilization buffer A (without urea) was ineffective at both pH 7.5 and pH 8, as the protein remained insoluble.

For the VP6 wild-type protein (Figure 4.13A), solubilization buffer B, which contained different concentrations of urea (2 M, 5 M, and 7 M), successfully solubilized the protein to different extents. At pH 7.5, 2 M and 5 M urea were only partially effective with some of the protein remaining insoluble in the pellet. However, with 7 M urea, the VP6 protein was fully solubilized. At pH 8, both 2 M and 5 M urea concentrations showed similar effectiveness as pH 7.5, partially solubilising the protein. Whereas with 7 M urea, the concentration achieved complete solubilization. These results suggest that the solubility of the VP6 wild-type protein expressed in *E. coli* NiCo21 (DE3) cells is highly dependent on higher concentrations of urea, particularly 7 M. Additionally, a higher pH (pH 8) appears to also enhance protein solubilization.

For the VP6 mutant protein (Figure 4.13B) solubilization buffer B containing different concentrations of urea (2 M, 5 M, and 7 M), was able to solubilize the protein to different degrees depending on the pH (pH 7.5 and pH 8). At pH 7.5, all tested urea concentrations (2 M, 5 M, and 7 M) successfully solubilized all the insoluble protein. Whereas at pH 8, urea concentrations of 2 M and 7 M partially solubilised the protein, with some of the protein remaining in the insoluble fraction. However, with 5 M urea, complete solubilization was achieved. Overall, the VP6 mutant protein showed successful solubilization across different pH levels and urea concentrations, although some conditions resulted in only partial solubilization.

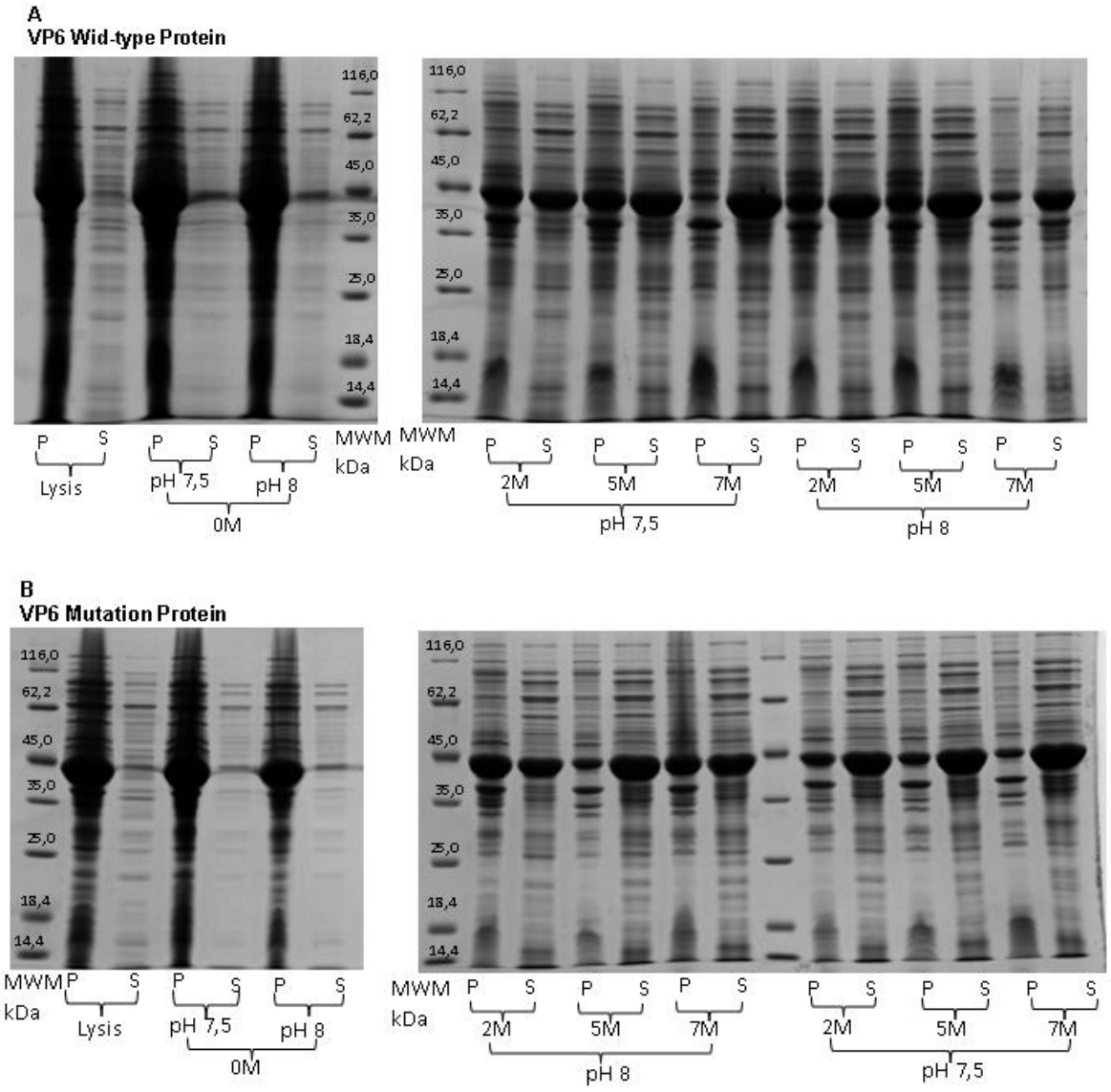


Figure 4.13: VP6 Wild-type and VP6 Mutant Protein Solubilisation

The figure shows the successful solubilisation of the protein. The overexpressed VP6 wild-type protein was successful at 7 M urea for both pH 7.5 and pH 8 while urea concentrations of 2 M and 5 M partially solubilised the protein **(A)**. The overexpressed VP6 mutant protein on the NiCo21 (DE3) was solubilised successfully at pH 7.5 for both urea concentrations and partially at pH 8 except for 5 M urea **(B)**. The solubility of the protein was compared among the lysed P (pellet), S (supernatant) and 0 M urea of the proteins alongside the molecular weight marker (MWM).

Protein solubility may be influenced by factors such as urea concentration, pH, amino acid sequence, and structural conformation (Trevino *et al.* 2007; Wang *et al.* 2024). The VP6 mutant protein was intentionally engineered with six amino acid substitutions, to improve solubility compared to the VP6 wild-type protein. Both proteins required urea for solubilisation, with 5 M and 7 M urea being most effective. The VP6 wild-type protein required a higher urea concentration, while the VP6 mutant protein solubilised completely at pH 7.5 with both 2 M, 5 M, and 7 M urea concentration. These findings underscore the importance of protein design and buffer conditions when optimizing solubilisation from inclusion bodies.

4.5 VP6 Protein Affinity Chromatography Purification

Affinity chromatography was used to selectively purify the solubilized VP6 protein. The binding affinity of the targeted protein was evaluated by quantifying the amount of target protein that effectively interacts/binds with the affinity resin (immobilised ligand) and enhancing binding efficiency conditions, such as pH and ionic strength. The purity of the eluate was assessed by monitoring the contaminants that co-elute alongside the target protein, with adjustments made to buffer conditions, including salt concentration, pH, or urea concentration, to reduce co-elution of non-target proteins/non-binding proteins and improve overall purity. Based on the solubilization results discussed in Section 4.4, urea concentrations of 5 M and 7 M at pH 7.5 and pH 8.0 were selected for purification trials to ensure maximal purification of the targeted protein. To further enhance protein purity, an additional step involving the isolation of inclusion bodies, (described in Section 4.3) was incorporated into selected trials. These purification experiments were conducted both with and without the inclusion body wash steps outlined in Section 4.3, allowing direct comparison of their impact on overall protein purity.

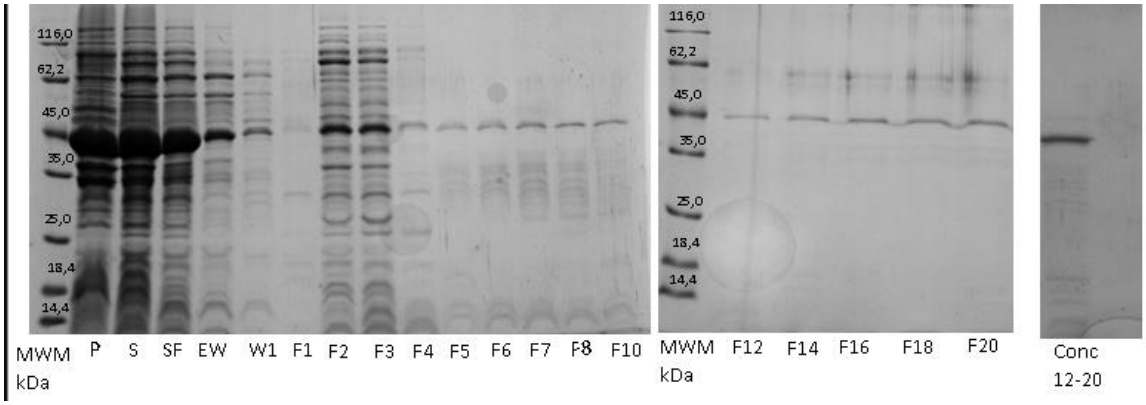
4.5.1 VP6 Wild-type Protein Purification

The purification of the VP6 wild-type protein was assessed through different trials, each aiming to resolve the specific challenges encountered in the previous attempts.

Trial 1 (Figure 4.14), the VP6 protein proved challenging to purify due to significant protein loss during the loading phase, where most of the VP6 protein passed through the His-tag column and was detected in the flow-through (labelled SF). The 300 mM NaCl was used as a standard for buffers throughout the purification steps, aiming to

reduce non-specific binding, stabilize the protein, and maintain solubility; however, it proved ineffective in removing non-specifically bound contaminants, as these remained on the column and continued to co-elute with the target protein. During the first elution gradient, the VP6 protein began to elute in small amounts at an imidazole concentration of approximately 159 mM, as indicated by an increased A_{280} absorbance (fractions F2–F3). Following this, a second elution gradient was applied, starting from the final concentration of the first gradient. In this second gradient, VP6 continued to elute across multiple fractions (F3–F18), although no distinct A_{280} peak was observed, suggesting a more gradual release of the protein. The extended elution of VP6 may be due to the presence of non-specifically bound proteins or strongly associated impurities co-eluting with the VP6 protein, leading to a gradual and spread-out elution profile. Even the fractions that appeared cleaner upon collection were still contaminated after concentration (F12 to F18).

Due to the poor VP6 protein's binding affinity observed initially, Trial 2 (Figure 4.15) focused on enhancing the protein's interaction with the His-tag column by minimizing imidazole in the binding buffer (5 mM) to allow strong, specific binding to the metal ions on the resin. This led to improved binding of VP6 during loading, decreasing the VP6 protein in the flow-through. Additionally, a wash step using 150 mM imidazole was added after protein binding to further reduce non-specific interactions and enhance purity prior to conducting the elution gradient. However, weakly bound VP6 still eluted alongside contaminants during the wash steps, and despite some improvements, the gradient elution again showed early A_{280} peaks mainly with impurities (F1 to F3). The VP6 protein began to elute at a slightly higher imidazole concentration, around 162.95 mM, but continued to co-elute with contaminants throughout the gradient. This showed a prolonged and impure profile that continued even after the lower A_{280} absorbance (F4 to F22), the contaminants persisted even after concentration (F5 to F22).



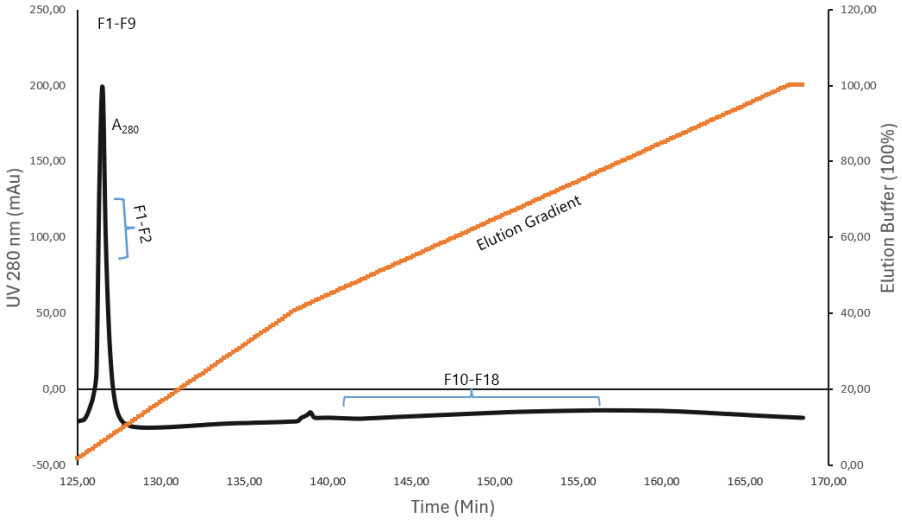


Figure 4.14: VP6 Wild-type Protein Purification Trial 1

A purification attempt was conducted for solubilised VP6 wild-type protein with equilibration buffer containing 0 mM imidazole and an elution buffer gradient with 500 mM imidazole. The corresponding figure displays the elution profile, where the black line represents absorbance at 280 nm (A_{280}), indicating protein concentration, while the orange line displays the linear imidazole gradient during elution. The gradient was established using a wash buffer containing 150 mM and an elution buffer with 500 mM imidazole. Fractions, labelled on the graph, were analysed using SDS-PAGE, which showed the flow-through, washes, and elution samples. The sample descriptions include P for pellet, S for supernatant (representing the original samples), FT for flow-through, S+E for an equilibrated sample before loading, EW for the equilibration wash post-loading, W1 for wash 1, F for the fractions collected during elution, MWM for the molecular weight marker, measured in kDa and "conc" for concentrated fractions.

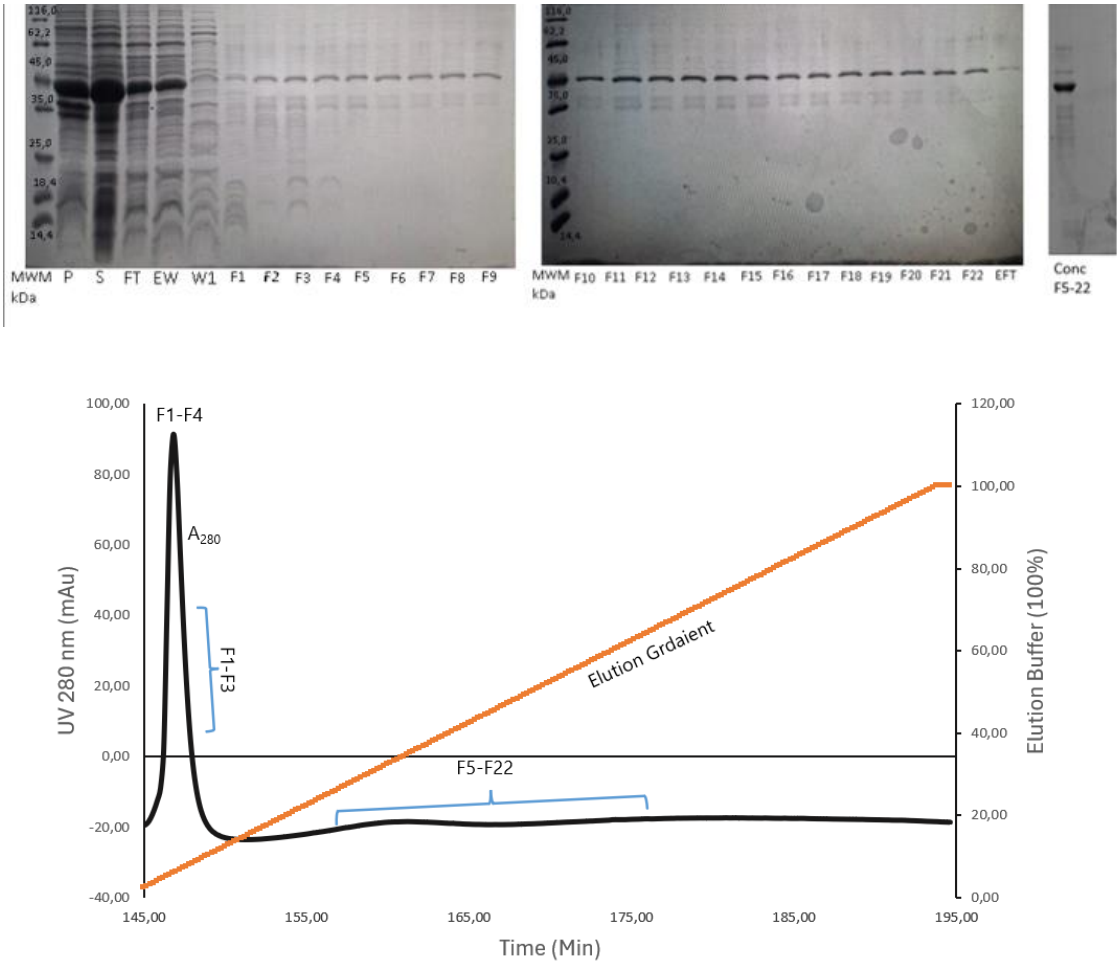


Figure 4.15: VP6 Wild-type Protein Purification Trial 2

A purification attempt was conducted for solubilised wild-type VP6 protein with equilibration buffer containing 5 mM imidazole and an elution buffer gradient with 500 mM imidazole. The corresponding figure displays the elution profile, where the black line represents absorbance at 280 nm (A_{280}), indicating protein concentration, while the orange line displays the linear imidazole gradient during elution. The gradient was established using a wash buffer containing 150 mM and an elution buffer with 500 mM imidazole. Fractions, labelled on the graph, were analysed using SDS-PAGE, which showed the flow-through, washes, and elution samples. The sample description includes P for pellet, S for supernatant (representing the original samples), FT for flow-through, EW for the equilibration wash post-loading, W1 for wash 1, F for the fractions collected during elution, MWM for the molecular weight marker, measured in kDa and "conc" for concentrated fractions.

To further eliminate the non-specific interactions, 50 mM imidazole was added to the binding buffer in Trial 3 (Figure 4.16), which significantly reduced the binding of unwanted proteins to the column, resulting in minimal loss of the protein during the loading phase. However, the pH was adjusted from 8.0 to 7.5 to maintain histidine in a deprotonated state for effective His-tag binding. This adjustment was also intended to improve protein stability and reduce aggregation. In addition, the NaCl concentration was lowered from 300 mM to 150 mM. Despite these modifications, persistent non-specific binding between contaminants and the resin or the target protein was still observed after the wash steps with 50 mM and 150 mM imidazole. Although the elution again began early at approximately 161.25 mM imidazole with a high A_{280} absorbance peak (F1 to F2), the protein continued to elute gradually throughout the gradient, though with a lower A_{280} signal, indicating a prolonged and less concentrated elution profile. Despite some fractions appearing cleaner upon collection (F30 to F38), VP6 protein remained present and impure with other proteins after concentration (F15-F28), indicating that non-specific binding proteins still interact with the column matrix, hindering the purification of the protein.

To improve purification, 7 M urea was used to denature VP6 and expose the His-tag. Imidazole (80 mM) was increased to reduce nonspecific binding to the Ni^{2+} resin and NaCl was lowered to 150 mM to enhance binding. A preliminary inclusion body wash step was also added to remove cell debris before column purification in Trial 4 (Figure 4.17). There was a reduction in the protein loss during the loading and wash phases, though a small amount of VP6 protein was lost with non-specifically bound proteins. Elution of the VP6 protein began earlier in the gradient, at about 128.36 mM imidazole, with a high A_{280} absorbance peak corresponding to the elution of the VP6 protein (F4 to F19), but the eluted protein still co-eluted with impurities. The VP6 protein continued to elute throughout the entire gradient although at low A_{280} absorbance peak (F20 to F41), the prolonged elution again pointed to weak/non-specific binding or co-purification of similarly binding contaminants. Despite all the VP6 protein washes (isolation of inclusion bodies wash step and the column washes), the fractions collected remained contaminated with reduced contaminants after concentration (F4 to 19, F20 to 31 and F31 to F41).

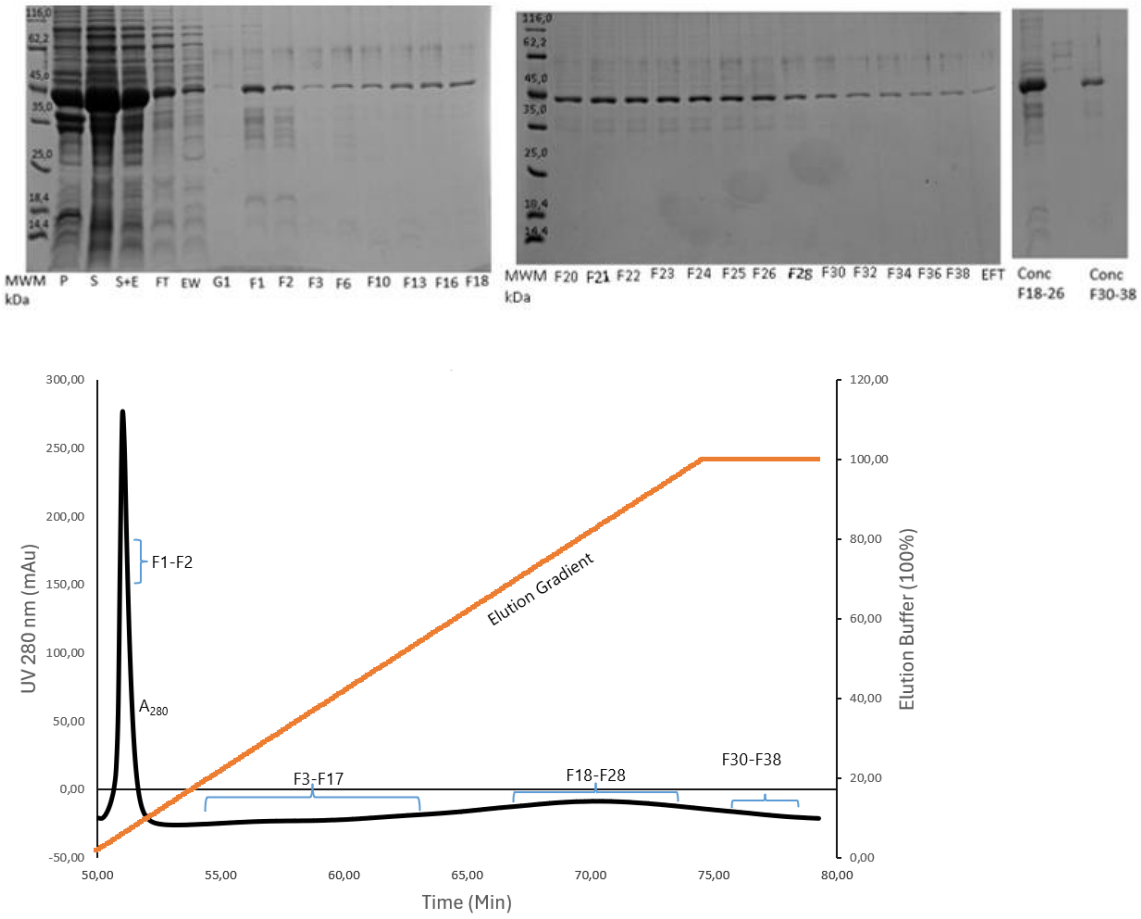


Figure 4.16: VP6 Wild-type Protein Purification Trial 3

A purification attempt was conducted for solubilised wild-type VP6 protein with equilibration buffer containing 50 mM imidazole and an elution buffer gradient with 600 mM imidazole. The corresponding figure displays the elution profile, where the black line represents absorbance at 280 nm (A_{280}), indicating protein concentration, while the orange line illustrates the linear imidazole gradient during elution. The gradient was established using a wash buffer containing 150 mM and an elution buffer with 600 mM imidazole. Fractions, labelled on the graph, were analysed using SDS-PAGE, which showed the flow-through, washes, and elution samples. The sample description includes P for pellet, S for supernatant (representing the original samples), S+E for an equilibrated sample before loading. FT for flow-through, EW for the equilibration wash post-loading, G1 for gradient wash from 50mM imidazole to 150 mM imidazole as a normal wash, F for the fractions collected during elution, MWM for the molecular weight marker, measured in kDa and "conc" for concentrated fractions.

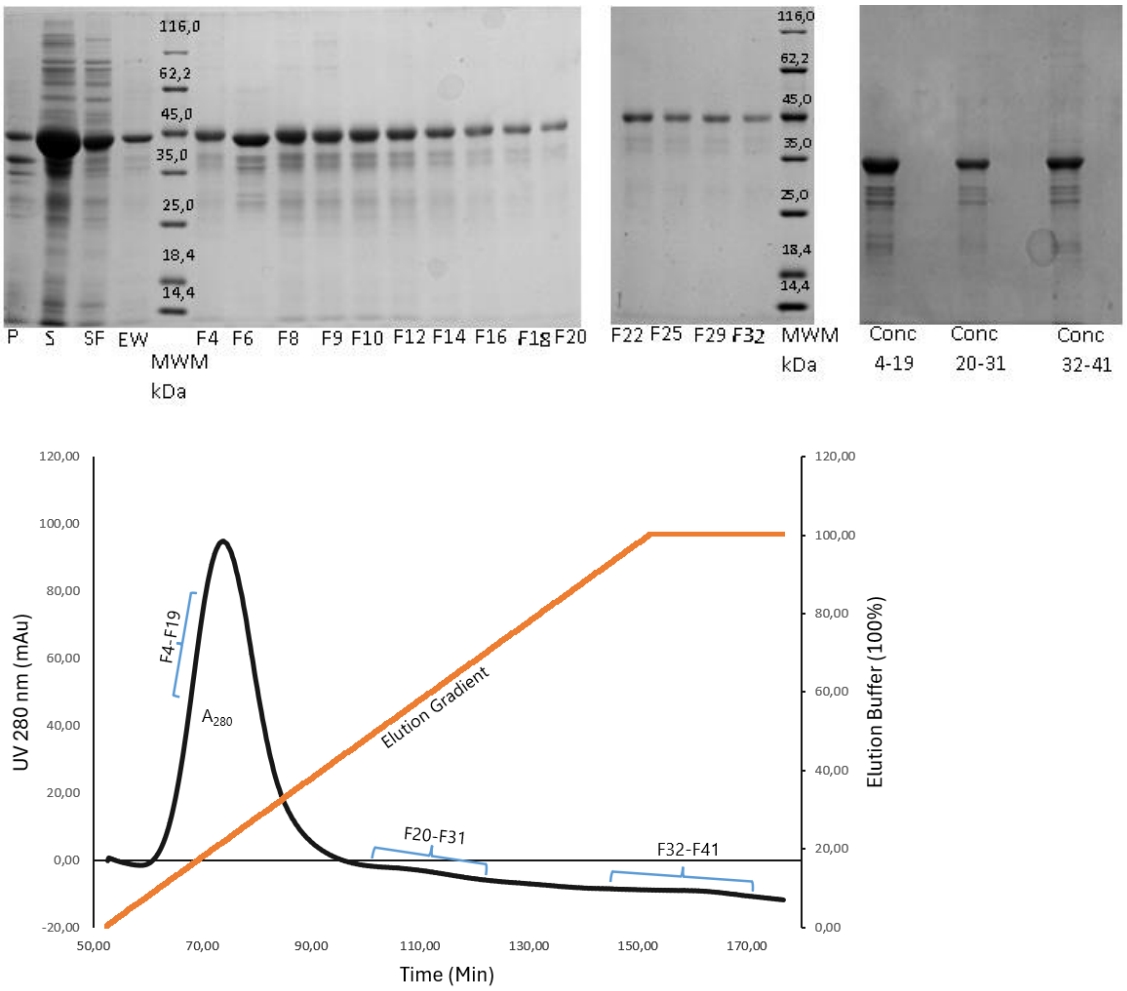


Figure 4.17: VP6 Wild-type Protein Purification Trial 4

A purification attempt was conducted for solubilised wild-type VP6 protein with equilibration buffer containing 80 mM imidazole and an elution buffer gradient with 600 mM imidazole. The corresponding figure displays the elution profile, where the black line represents absorbance at 280 nm (A_{280}), indicating protein concentration, while the orange line indicates the linear imidazole gradient during elution. The gradient was established using the equilibration buffer containing 80 mM and an elution buffer with 600 mM imidazole. Fractions, labelled on the graph, were analysed using SDS-PAGE, which showed the flow-through, washes, and elution samples. The sample description included P for pellet, S for supernatant (representing the original samples), FT for flow-through, EW for the equilibration wash post-loading, F for the fractions collected during elution, MWM for the molecular weight marker, measured in kDa and "conc" for concentrated fractions.

During the final trial (Figure 4.18), 0.5% Tween-20 was incorporated into the wash buffer with the NaCl concentration increased from 150 mM to 300 mM. This aimed to disrupt weak hydrophobic and electrostatic interactions between VP6, weakly bound contaminants, and the column (Han *et al.* 2019). A small amount of the VP6 was lost during the loading phase and the column wash step, this approach helped reduce a significant amount of the non-specific protein binding. The equilibration wash on the VP6 protein did not assist in removing any contaminants but was used to remove the 0.5% Tween-20 from the column to ensure the purity of the protein. Elution of the VP6 protein began at the beginning of the elution gradient with a comparable imidazole concentration of 139.8 mM, with a high A_{280} absorbance yet the target protein still eluted alongside contaminants (F13 to F27) with a reduced number and concentration of the contaminants. There was a slight shift in the peak where a second peak appeared later in the gradient, indicating continued co-elution of VP6 with additional impurities (F28 to F41). The VP6 did elute impurely throughout the elution gradient despite the additional wash step (F13 to 27, F28 to 41, F42 to F50,) whereas some of the protein fractions showed contaminants after concentration (F51 to F60 and F61 to F70) also some protein contaminants differ per fractions where F21, F33 and F35 were not included during concentration due to more concentrated contamination bound. While the current protocol used a single elution gradient, the observation of multiple peaks and different contaminant profiles raises the possibility that extending or optimizing the gradient might improve resolution between the target protein and contaminants. Although this was not tested in the present run, it would be a logical next step to assess whether the two peaks are indeed separable under modified gradient conditions.

Purifying VP6 wild-type protein presented ongoing challenges despite multiple protocol adjustments. In Trial 1, protein loss and contaminants co-elution indicated poor binding and ineffective separation. Trial 2 improved binding with altered imidazole levels but failed to prevent impurity co-elution. Trial 3 adjusted buffer pH and NaCl to reduce non-specific interactions, which minimized loss but didn't eliminate contaminants. Trial 4 used 7 M urea and an inclusion body wash step, yet impurities persisted. Finally, Trial 5 introduced 0.5% Tween-20 and high NaCl, reducing weak interactions.

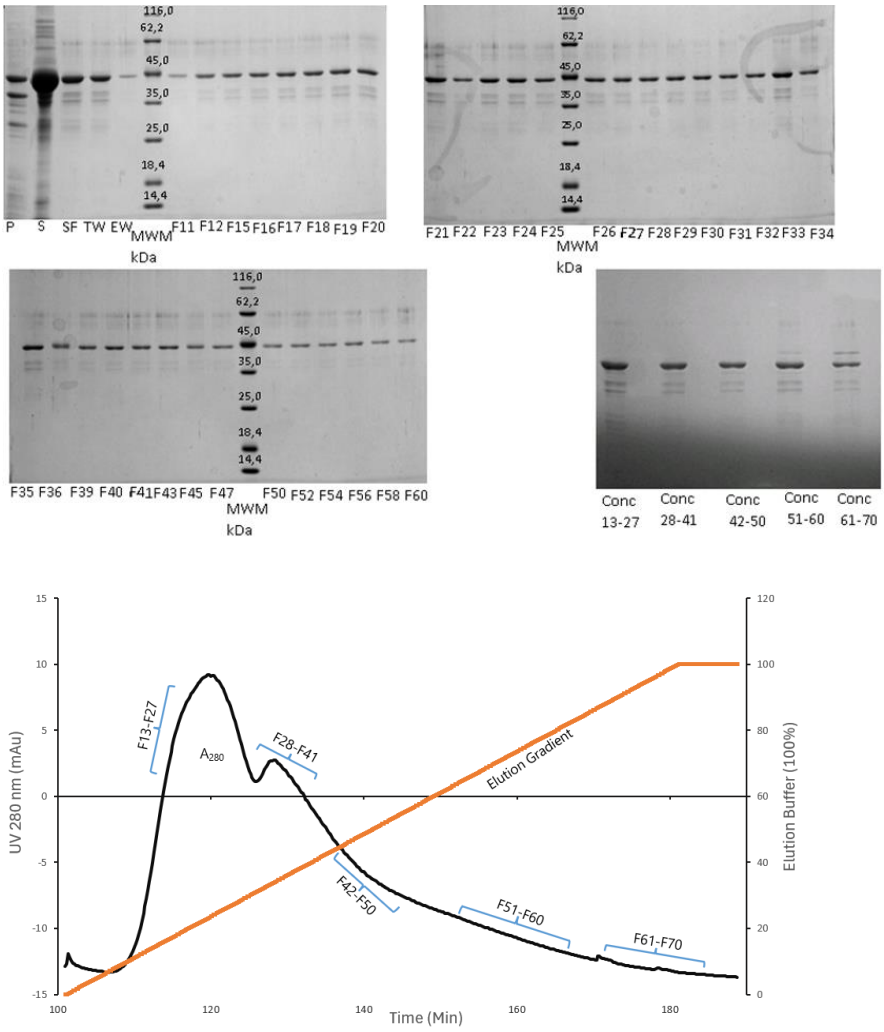


Figure 4.18: VP6 Wild-type Protein Purification Trial 5

A purification attempt was conducted for solubilised wild-type VP6 protein with equilibration buffer containing 80 mM imidazole and an elution buffer gradient with 600 mM imidazole. The corresponding figure displays the elution profile, where the blue line represents absorbance at 280 nm (A_{280}), indicating protein concentration, while the red line indicated the linear imidazole gradient during elution. The gradient was established using an equilibration buffer containing 80 mM imidazole and an elution buffer with 600 mM imidazole. Fractions, labelled on the graph, were analysed using SDS-PAGE, which showed the flow-through, washes, and elution samples. The sample description included P for pellet, S for supernatant (representing the original samples), FT for flow-through, TW for 0.5% Tween-20 wash, EW for the equilibration wash post-loading, F for the fractions collected during elution, MWM for the molecular weight marker, measured in kDa and "conc" for concentrated fractions.

Despite some improvements, complete purification remained partial due to remaining co-eluted impurities. While this helped decrease weak non-specific binding, VP6 continued to co-elute with some bands that were less concentrated, making the purification process partially successful, with reduced but not eliminated impurities.

4.5.2 VP6 Mutant Protein Purification

The purification of the VP6 mutant protein was based on the optimized conditions established for the VP6 wild-type protein, where contamination had previously persisted during binding and elution. Similar conditions were initially applied to the VP6 mutant protein to assess whether they would yield comparable purification efficiency. Different conditions, as per Trial 4 were implemented to improve the protein binding and purity of the VP6 mutant protein (Trial 1: Figure 4.19). These conditions facilitated better column interaction, although a small amount of the VP6 protein was lost during the loading and equilibration wash phases. As the elution gradient began (80 mM–600 mM imidazole), VP6 protein began to elute at 123.68 mM with a high A_{280} peak, suggesting a substantial protein yield. However, this peak co-eluted with numerous impurities (F4 to F16), followed by a second shift in a peak (F17 to F27). However, the level of impurities remained largely bound, continuing to co-elute with the target protein. The VP6 elution continued throughout the gradient with lower A_{280} absorbance (fractions F28 to F48), still eluting detectable impurities. These results suggested that the VP6 mutant protein exhibited similar purification challenges as the VP6 wild-type protein. Tightly bound contaminants or protein remain associated with the VP6 mutant protein, altering its binding behaviour and resulting in delayed, prolonged elution across the gradient. As with the VP6 wild-type protein, the collected fractions (F1–16, F17–27, and F28–48) remained contaminated after concentration, indicating that the mutation did not significantly enhance the purity of the final VP6 protein.

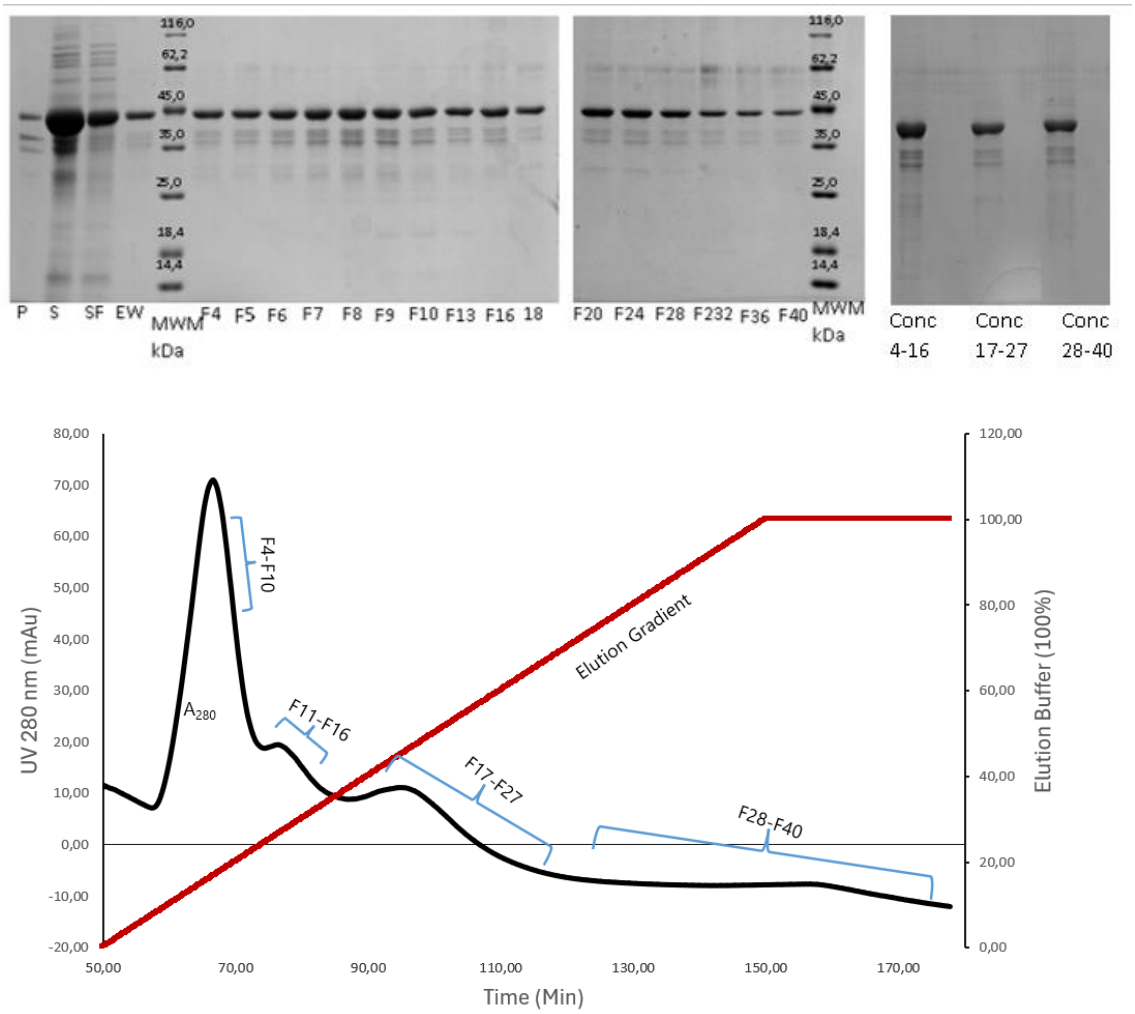


Figure 4.19: VP6 Mutant Protein Purification Trial 1

A purification attempt was conducted for solubilised VP6 mutant protein with equilibration buffer with 80 mM imidazole and an elution buffer gradient with 600 mM imidazole. The corresponding figure displays the elution profile, where the black line represents absorbance at 280 nm (A_{280}), indicating protein concentration, while the red line indicates the linear imidazole gradient during elution. The gradient was established using the equilibration buffer containing 80 mM and an elution buffer with 600 mM imidazole. Fractions labelled on the graph were analysed using SDS-PAGE, which showed the flow-through, washes, and elution samples. The sample description included P for pellet, S for supernatant (representing the original samples), FT for flow-through, EW for the equilibration wash post-loading, F for the fractions collected during elution, MWM for the molecular weight marker, measured in kDa and "conc" for concentrated fractions.

The second trial (Figure 4.20) was implemented to enhance purity by introducing a 0.5% Tween-20 wash, aiming to disrupt weak hydrophobic interactions and reduce non-specific binding to both the resin and VP6 protein. Even though the binding affinity of the protein improved with a small amount of protein lost during the loading phase, small amounts of VP6 protein and contaminants were still observed in the 0.5%Tween-20 and equilibration wash steps. The strategy showed limited success in removing strongly bound contaminants. During the elution gradient, elution began at a slightly higher concentration of 136.68 mM with a higher A_{280} absorbance peak (F21 to F30), the A_{280} peak shifted further into the gradient, eluting the targeted VP6 with fewer impurities. VP6 continued to elute across the gradient at lower A_{280} absorbance (fractions F31 to F50), still co-eluting with impurities. As a result, VP6 remained contaminated throughout the elution. Despite the wash steps, the concentrated fractions of the VP6 protein (F8 to F20, F21 to F30, F31 to F40, and F41 to F50) remained impure.

In the final trial (Figure 4.21), the focus remained on improving the purity of VP6 by eliminating all the co-eluting contaminants. The combination of 80 mM imidazole, a repeated 0.5% Tween-20 wash, and an increased NaCl concentration of 300 mM was applied to further disrupt non-specific interactions. During the loading phase, a small amount of the VP6 was lost through the flow-through during the loading phase and the washes. As in the previous trial, the Tween-20 wash targeted weakly interacting proteins, while the equilibration washes mainly served to remove residual detergent, ensuring that the elution phase was unaffected. Despite these improvements, during the elution gradient with a higher A_{280} absorbance peak, VP6 eluted at 138.76 mM imidazole, which was almost consistent with the previous trial. Although some improvements were noted in reducing the overall level of contaminants, fractions (F18 to F64) showed visible impurities, while fractions (F65 to F80) showed contaminants after concentration. Following concentration of the fractions (F18 to F31, F32 to F64, and F65 to F80), residual impurities remained present, though at reduced levels Thus, although contaminant levels may have decreased slightly, significant purification was not achieved. Interestingly, although Han *et al.* (2019) showed that 250 mM NaCl effectively reduced non-specific binding, this study found 300 mM NaCl ineffective, indicating that the influence of salt concentration on purification efficiency is highly protein- or system-specific.

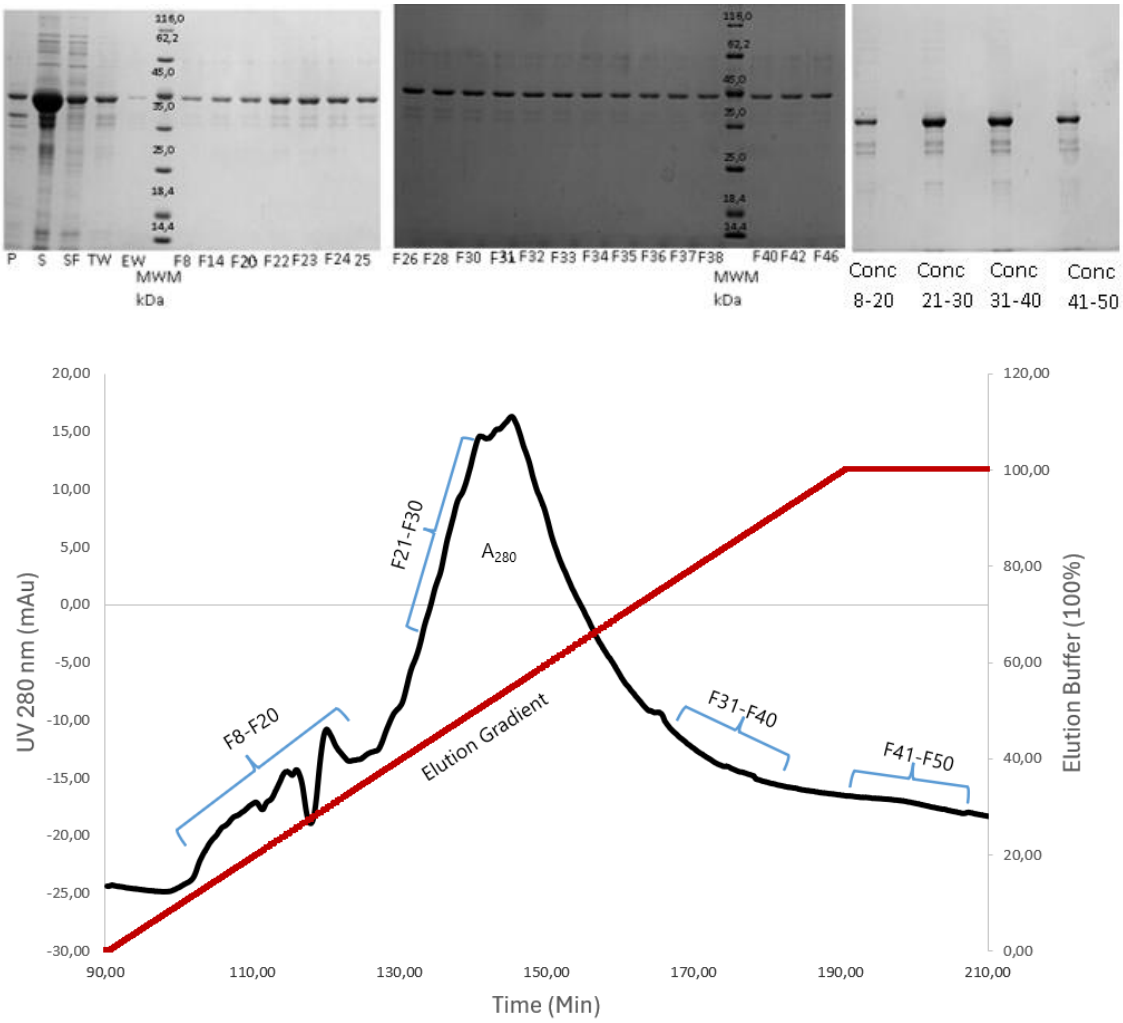


Figure 4.20: VP6 Mutant Protein Purification Trial 2

A purification attempt was conducted for solubilised VP6 mutant protein with equilibration buffer with 80 mM imidazole and an elution buffer gradient with 600 mM imidazole. The corresponding figure displays the elution profile, where the black line represents absorbance at 280 nm (A_{280}), indicating protein concentration, while the red line displayed the linear imidazole gradient during elution. The gradient was established using a wash buffer containing 80 mM and an elution buffer with 600 mM imidazole. Fractions, labelled on the graph, were analysed using SDS-PAGE, which showed the flow-through, washes, and elution samples. The sample description included P for pellet, S for supernatant (representing the original samples), FT for flow-through, TW for 0.5% Tween-20 wash, EW for the equilibration wash post-loading, F for the fractions collected during elution, MWM for the molecular weight marker, measured in kDa and "conc" for concentrated fractions.

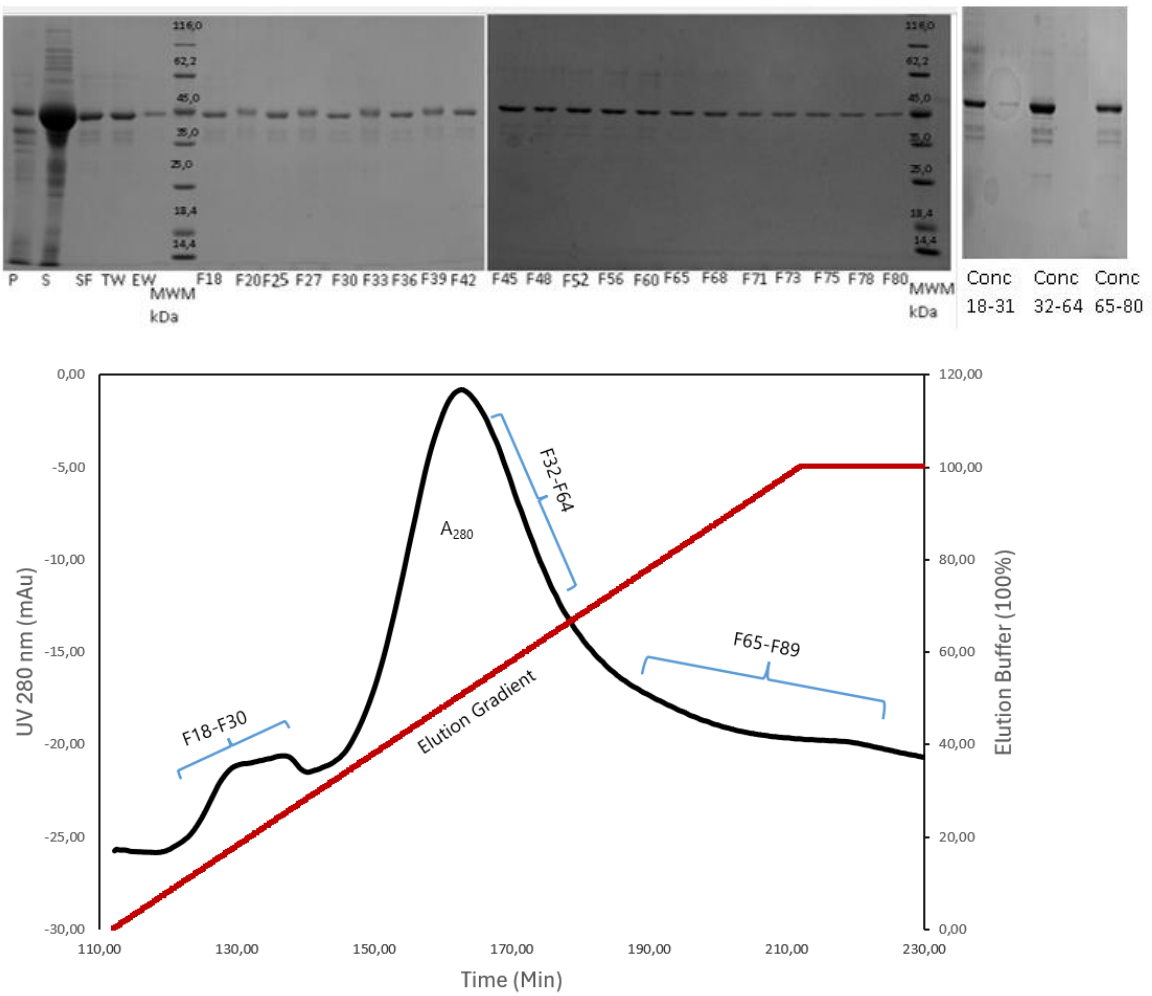


Figure 4.21: VP6 Mutant Protein Purification Trial 3

A purification attempt was conducted for solubilised wild-type VP6 protein with equilibration buffer containing 80 mM imidazole and an elution buffer gradient with 600 mM imidazole. The corresponding figure displays the elution profile, where the blue line represents absorbance at 280 nm (A_{280}), indicating protein concentration, while the red line indicated the linear imidazole gradient during elution. The gradient was established using an equilibration buffer containing 80 mM imidazole and an elution buffer with 600 mM imidazole. Fractions, labelled on the graph, were analysed using SDS-PAGE, which showed the flow-through, washes, and elution samples. The sample description included P for pellet, S for supernatant (representing the original samples), FT for flow-through, TW for 0.5% Tween-20 wash, EW for the equilibration wash post-loading, F for the fractions collected during elution, MWM for the molecular weight marker, measured in kDa and "conc" for concentrated fractions.

The purification of the VP6 mutant protein across three trials proved challenging to achieve high purity despite different modifications to the protocol. Initially, the inclusion of 7M urea and 80 mM imidazole improved His-tag exposure and minimized non-specific binding; however, significant co-elution of contaminants remained throughout the gradient. The second trial included 0.5% Tween-20 wash, improving protein binding and shifted the VP6 elution peak further along the gradient, but did not eliminate impurities. Finally, the third trial combined the previous strategies with increased salt concentration (300 mM NaCl), aiming to disrupt stronger protein-protein and hydrophobic interactions. Although this led to a slight reduction in contaminants, complete purification remained a challenge, with co-elution persisting even after concentration.

Overall, the purification of both VP6 proteins remained a challenge, primarily due to persistent co-elution with contaminants and early elution during the imidazole gradient. While binding efficiency was moderately improved through buffer optimization, such as the use of urea, different imidazole concentrations, NaCl adjustments, and Tween-20 washes, none of the strategies fully resolved the issue of contamination. The VP6 wild-type protein trials initially showed high protein loss during loading, which was reduced with further optimization of imidazole concentration in the binding buffer in the first three trials. However, despite all the optimisation, the target protein consistently co-eluted with impurities across all trials. Similarly, the VP6 mutant protein was purified under the same optimized conditions as the VP6 wild-type protein, including the use of 7 M urea, 80 mM imidazole, and reduced NaCl concentration and 0.5% tween-20. While the mutant showed slightly improved binding behaviour, it still exhibited similar challenges with co-elution of contaminants. Overall, the purification process for both the VP6 wild-type protein and VP6 mutant protein achieved only partial success, with reduced but not fully eliminated impurities.

4.6 VP6 Protein Concentration Determination

The absorbance for the VP6 wild-type protein and VP6 mutant protein were measured using an Applied Photophysics Chirascan Plus, DichroWeb (Leatherhead, United Kingdom) at a wavelength between 220 nm to 280 nm, to assess its structural characteristics based on wavelength-dependent spectral changes.

For both VP6 wild-type protein ● (Figure 4.22A) and VP6 mutant protein ● (Figure 4.22B) the absorbance confirmed the presence of the VP6 protein in 4 M urea. VP6 wild-type protein and VP6 mutant protein showed an absorbance peak at 280 nm and an increased absorbance signal again at 240 nm. The 280 nm peak indicated the presence of aromatic residues like tryptophan and tyrosine, as per Pace *et al.* (1986) who mentioned that UV absorbance above 275 nm increases due to the presence of tryptophan, tyrosine, and cystine (disulfide bonds) whereas absorbance at 220 nm assesses peptide bonds and secondary structure complementing the presence of aromatic signals at 280 nm. The 220 nm absorbance reflects a well-ordered secondary structure (alpha-helices and beta-sheets), confirming proper folding and structural integrity. The higher absorbance values observed in the VP6 wild-type protein displayed a higher protein concentration compared to the VP6 mutant protein. These findings support the successful solubilization and structural integrity of both proteins in 4 M urea, with the VP6 wild-type protein displaying stronger signals, indicating abundant or more preserved folding. Another important factor was the baseline scattering around 340 nm, which was higher for the VP6 wild-type protein than in the VP6 mutant protein, indicating larger aggregates or oligomers. In contrast, the VP6 mutant protein showed reduced scattering, suggesting lower aggregation. These spectral differences imply improved solubility and structural stability of the VP6 mutant protein under denaturing conditions despite lower overall concentration.

Figure 4.22C showed the concentration of the VP6 wild-type protein ● and VP6 mutant protein ● in 4 M urea concentration, determined using the Bradford Assay standard curve. For the VP6 wild-type protein, the concentration was 0.9837 mg/ml, and it was calculated to be 21.86 µM in comparison to the BSA concentration which was 0.9546 mg/ml, and it was calculated to be 21.21 µM as determined by the Bradford assay as a standard. For the VP6 mutation protein, the concentration was 0.8512 mg/ml, and it was calculated to be 18.92 µM in comparison to the BSA concentration which was 0.774 and it was calculated to be 17.2 µM as determined by the Bradford assay as a standard. The results of both VP6 protein were comparable to those obtained using the spectrophotometer and the Applied Photophysics Chirascan Plus. The yields for the VP6 wild-type protein were estimated as 3.94 mg and for VP6 mutant protein was 3.41 mg from 500 ml culture.

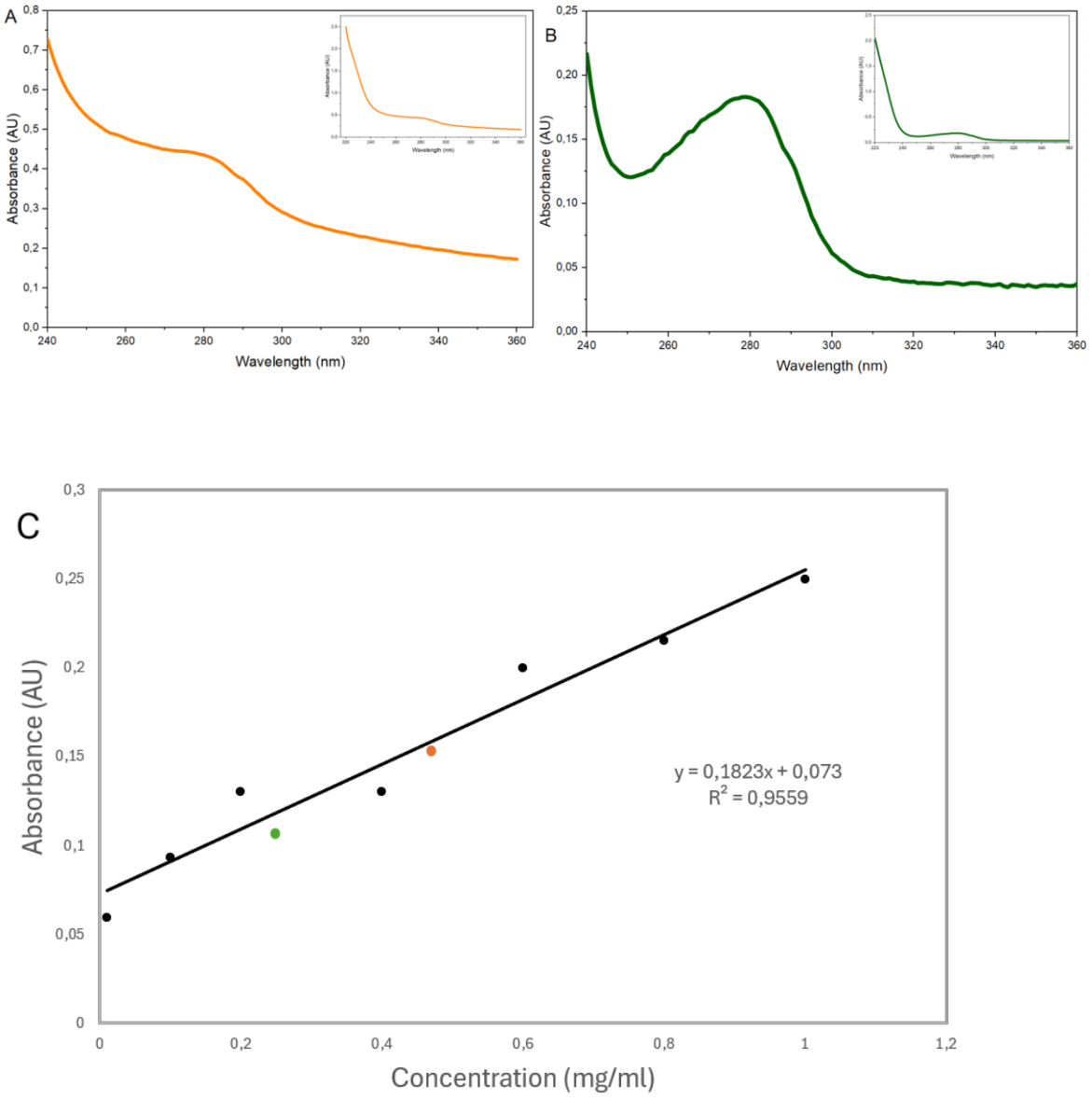


Figure 4.22: Purified VP6 Protein Absorbance Spectrum

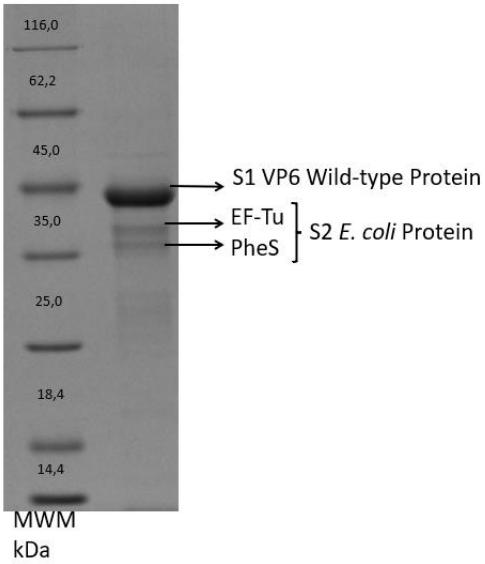
Shows the purified VP6 wild-type protein highlighted in orange **(A)** and VP6 mutant protein highlighted in green **(B),** absorbance measured between wavelengths of 360 nm to 240 nm with 4 M urea concentration. **(C)** Represent the standard curve of the BSA stock with a trendline equation used to calculate the VP6 protein concentration using the Bradford assay measured at 595 nm.

4.7 Mass Spectroscopy Analysis

Following the IMAC purification trials described in Section 4.5, both the VP6 wild-type and mutant proteins were eluted along with residual impurities. To confirm the identity of the VP6 protein and the contaminating protein bands (Figures 4.23 and Figure 4.24), a mass spectrometry-based approach was used. Peptides generated through trypsin and chymotrypsin digestion were separated by liquid chromatography (LC) and analysed using mass spectrometry. The resulting peptide sequences were compared against the UniProt database (Swiss-Prot and TrEMBL), including the supplied VP6 sequences and *E. coli* protein sequences from the expression system, to identify both target proteins and co-eluting contaminants.

Affinity purification of the VP6 wild-type protein using metal-affinity chromatography resulted in a prominent protein band, which was confirmed as the wild-type VP6 protein by trypsin digestion and LC-MS/MS analysis (Figure 4.23A). The 18 matching peptide sequences ● were identified, covering 53.4% of the protein sequence (Figure 4.23B). Alongside the primary band, two smaller, faint bands were observed and individually identified as Elongation Factor Tu (EF-Tu) and Phenylalanine-tRNA ligase alpha subunit (PheS), both proteins originating from *E. coli* (strain K-12) (Figure 4.23A). These proteins likely co-eluted with the target protein due to their high abundance in *E. coli* and potential nonspecific interactions with the metal affinity resin due to their hydrophobicity and sizes of 43 kDa and 40 kDa respectively. Such interactions may be facilitated by surface-exposed histidine residues or hydrophobic patches, which allow weak binding to the IMAC resin despite not having a His-tag (Robichon *et al.* 2011). Their molecular weights also fall within a similar range to VP6 protein, possibly contributing to co-purification under the chosen elution conditions. In the VP6 wild-type protein, methionine residues at positions 97, 100, 295, 300, 365, and 394 were identified as oxidized based on mass spectrometry analysis. The presence of multiple oxidized methionine, particularly clustered in the C-terminal region, suggested that the sites may be solvent-exposed and thus more prone to oxidative modification. This clustering could potentially influence the protein's folding or structural stability.

A



B

>VP6 Wild-type Sequence/1-397

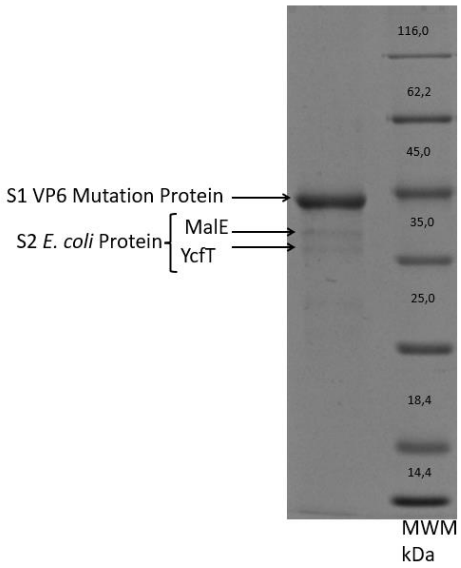
MEVLYSLSKTLKDARDKIVEGTLYSNVSDLIQQFNQMIVTMNGNDFQTGGIGNLPIRNWTFD
FGLLGTTLLNLDANYVETARTTIEYFIDFIDNVCMDEMARESQRNGVAPQSEALRKLAGIKF
KRINFNNSSEYIENWNLQNRRQRTGFVFHKPNIFPYSASFTLNRSQPMHDNLMGTMWLNAGS
EIQVAGFDYSCALNAPANIQQFEHIVQLRRALTTATITLLPDAERFSFPRVINSADGATTWF
FNPIILRPNNVEVEFLLNGQIINTYQARFGTIVARNFDTIRLSFQLMRPPNMTPAVNALFPQ
AQPFQHHATVGLTLRIESAVCESVLADANETLLANVTAVRQEYAIPVGPVFPPGMNWTELIT
NYSPSREDNLQRVFTVASIRSMLIK

Figure 4.23: Mass Spectrometry Analysis of VP6 Wild-type Protein

Shows the results of mass spectrometry analysis from the LCMS. (**A**) SDS-PAGE was used to represent the purification results, with MWM representing the molecular weight marker in kDa. Two samples were analysed sample 1 representing the VP6 Wild-type protein and sample 2 representing the *E. coli* protein K-12 (Elongation Factor Tu (EF-Tu) and Phenylalanine-tRNA ligase alpha subunit (PheS)). **(B)** shows the sequence coverage of 53.4% of the VP6 protein.

VP6 mutant protein purification yielded a strong prominent band which was confirmed as the target protein using chymotrypsin digestion and mass spectrometry analysis (Figure 4.24A). A total of 18 peptide sequences were matched to VP6 mutant protein ● with 42.8% sequence coverage. Importantly, mass spectrometry also confirmed the presence of some of the intended mutations. Out of the six engineered mutations, only two were confirmed/covered, amino acid positions A39 and A151 ● (Figure 4.24B), indicating partial but specific confirmation of the designed sequence changes. In addition to the VP6 band, the two small faint bands were identified as *E. coli* (strain K12) proteins: the Inner membrane protein YcfT and Maltose/maltodextrin-binding periplasmic protein (MalE). These proteins were confirmed as non-specifically binding due to similar physicochemical characteristics to VP6 protein such as hydrodynamic properties, molecular weights/sizes (39.8 kDa and 42.5 kDa, respectively), and possibly surface charge characteristics. These factors may cause them to interact similarly to the VP6 protein under the same purification conditions. Such non-specific binding by native *E. coli* proteins has been previously reported in IMAC workflows (Bolanos-Garcia and Davies, 2006). Furthermore, several methionine residues within the VP6 mutant protein were identified as prone to oxidation, specifically at positions 37, 41, 172, 177, 180, and 394. Oxidation at these sites might influence how the VP6 mutant protein functions, folds, stabilize, or assemble into higher-order structures (Hu *et al.* 2008). Notably, position 394 appeared in both wild-type and mutant sequences, indicating a conserved oxidation hotspot with potential structural relevance or regulatory importance. Oxidation at this site might influence how the protein folds or maintains stability, potentially affecting its overall 3D shape, and with regulatory importance**,** oxidation could also act as a molecular switch that modulates protein function, interactions, or degradation pathways (Kim *et al.* 2013).

A



B

> VP6 Mutant Sequence /1-397

MEVLYSLSKTLKDARDKIVEGTLYSNVSDLIQQFNQMIATMNGNDFQTGGIGNLPIRNWTFD
FGLLGTTLLNLDANYVETARTTIEYFIDFIDNVCMDEMARESQRNGLAPQSEALRKLAGIKF
KRINFNNSSEYIENWNLQNRRQRTGFTFHKPNIFPYSASFTLNRSQPMHDNLMGTMWLNAGS
EIQVAGFDYSCALNAPANIQQFEHIVQLRRALTTATITLLPDAERFSFPRVINSADGATTWH
FNPTILRPNNVEVEFLLNGQIINTYQARFGTITARNFDTIRLSFQLMRPPNMTPAVNALFPQ
AQPFQHHATVGLTLRIESAVCESVLADANETLLANVTAVRQEYAIPVGPVFPPGMNWTELIT
NYSPSREDNLQRVFTVASIRSMLIK

Figure 4.24: Mass Spectrometry Analysis of VP6 Mutant Protein

Shows the results from the mass spectrometry analysis using the LCMS. **(A)** SDS-PAGE was used to represent the purification results, with MWM representing Molecular weight marker in kDa. Two samples were analysed sample 1 representing the VP6 mutant protein and sample 2 representing *E. coli* protein K-12 (Inner membrane protein YcfT and Maltose/maltodextrin-binding periplasmic protein (MalE). **(B)** shows the sequence coverage of 42.3% of the VP6 protein with two mutants covered position A39 and position A151.

4.8 Characterisation of the Secondary Structure

Far-UV circular dichroism (CD) was used to characterize the secondary structure of both VP6 wild-type protein and VP6 mutant protein. Far-UV circular dichroism spectra measurements were taken between 190 nm and 250 nm to identify the key secondary structural features. Alpha-helical contents were indicated by a positive peak near 190 nm to 195 nm and with a negative trough at 208 nm, 218 nm, and 222 nm, in contrast, the beta-sheet structures were indicated by a trough at 215 nm and a peak near 198 nm (Li *et al.* 2000; Wei *et al.* 2014; Kelly *et al.* 2005). The ATR–FTIR spectroscopy was used to gain deeper insight into the secondary structural characteristics as a complementary step to Far UV circular dichroism by analysing absorbance in the mid-infrared range (525 cm^{-1}–4000 cm^{-1}), with particular attention to the amide I (1650 cm^{-1}) and amide II (1540 cm^{-1}) bands which are key indicators of protein secondary structures (Krimm and Bandekar, 1986; Miller *et al.* 2013). Different urea concentrations (0 M, 4 M, 5 M, and 7 M) were used to analyse the structural changes in both VP6 wild-type protein and VP6 mutant protein.

4.8.1 VP6 Wild-type Protein Secondary Structure

For the VP6 wild-type protein (Figure 4.25), Far-UV circular dichroism spectrum and ATR-FTIR spectrum were measured under different denaturant concentrations (0 M, 4 M, 5 M, and 7 M) to analyse the secondary structure compositions, which includes alpha-helices, beta-sheets, turns, and random coils. These complementary techniques allowed investigation into the unfolding behaviour and conformational changes of VP6 protein in response to chemical denaturation. Far-UV circular dichroism spectra of the VP6 wild-type protein revealed distinct changes in secondary structure content influenced by urea concentration (Figure 4.25A). At 0 M urea, the Far-UV circular dichroism spectrum showed a strong signal near 205 nm, which might be alpha-helices or disordered protein as it was close to the characteristic 208 nm wavelength minimum associated with alpha-helical content. However, the absence of defined spectrum at both 218 nm and 222 nm suggested partial structural loss or the presence of disordered regions due to formation of visible aggregation.

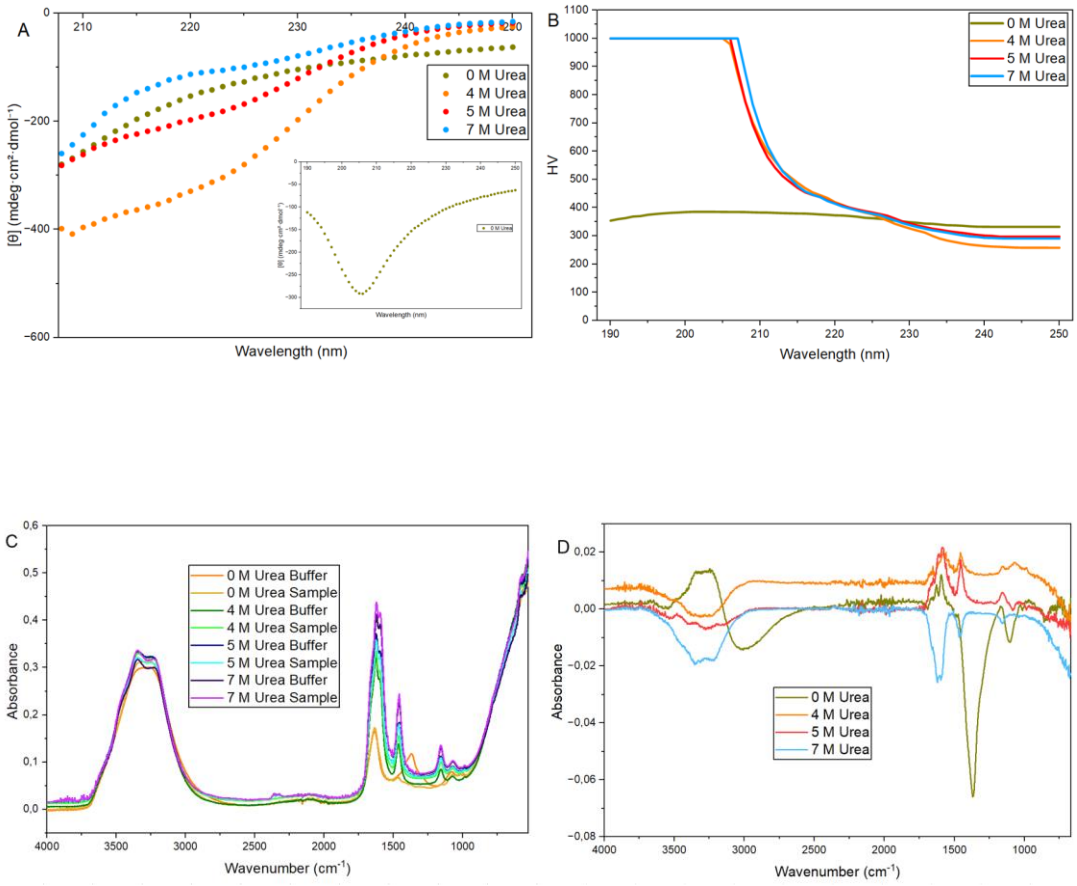


Figure 4.25: VP6 Wild-type Protein Secondary Structure Characterisation

The VP6 wild-type protein Far-UV circular dichroism spectra were measured between the wavelengths of 250 nm to 190 nm. Then protein was measured in urea concentrations of 0 M highlighted in brown,4 M highlighted in orange, 5 M highlighted in red, and 7 M highlighted in blue **(A)** and the HV using 6 µM **(B)**. The secondary structure was further confirmed using ATR–FTIR absorbance spectrum between the wavenumber of 4000 cm^{-1} to 500 cm^{-1} of the VP6 wild-type protein analysed using similar urea concentrations of 0 M, 4 M, 5 M, and 7 M. **(C)** Shows overall measurement of the samples and buffer. **(D)** Shows the spectrum highlights of the protein-specific absorbance peaks more clearly.

Similar observations were reported by Russel and Gildenhuys, (2018), who used 5 M urea to successfully analyse the bluetongue virus VP7 protein, highlighting the importance of urea in preventing aggregation during structural studies or analysis. At 4 M urea concentration, characteristic negative ellipticity were observed at key wavelength (218 nm and 222 nm) representing the presence of the secondary structure under mild denaturation. The presence of these signals suggests that alpha-helices and beta-sheet content remained partially intact, possibly due to the stabilizing effect of the mild denaturant preventing extensive aggregation. At 5 M urea, a reduction in ellipticity at 222 nm was observed, indicating unfolding and loss of alpha-helical content. A small residual signal at 218 nm suggested the presence of antiparallel beta-sheets, potentially arising from misfolded or reorganized structures. This transition is an indicator of partial denaturation and conformational rearrangement. At 7 M urea concentration, the Far-UV circular dichroism spectrum substantially decreased in ellipticity at both 222 nm and 218 nm, confirming the complete loss of secondary structure and full protein unfolding. The loss of defined secondary structural features at this concentration corroborated the denaturing effect of high urea concentration, leading to structural destabilisation.

Figure 4.25B showed the analysis of the HV to confirm signal to noise across the spectral range. For 0 M urea concentrations, data between 250 nm and 190 nm remained within acceptable HV thresholds. The data between 212 nm to 190 nm for all urea concentrations was deemed unreliable as per Miles and Wallace, (2016) and Siligardi *et al.* 2018 who confirmed that high tension voltage (HV) above 600 V indicating insufficient light intensity, an increase in noise making the Far-UV circular dichroism signal unreliable, whereas data below 600 V is deemed reliable. In contrast, the HV values between 250 nm to 213 nm were below 600 V, producing reliable data, high-quality spectral data for all the urea concentrations of the protein. Therefore, the structural interpretation of the alpha-helix or beta-sheet content was based primarily on the wavelength between 250 nm to 213 nm region, where the signal quality was acceptable. Overall, Far-UV circular dichroism analysis demonstrated that VP6 wild-type protein exhibits limited structural stability under native conditions and undergoes progressive unfolding with increasing urea concentrations. The improved structural definition at 4 M urea suggests a stabilizing effect at moderate denaturant concentrations, while higher concentrations (5–7 M) led to structural loss.

ATR–FTIR spectroscopy analysis was additionally performed to validate the presence of secondary structural elements in the VP6 wild-type protein (Figure 4.25C), showing the absorbance spectra of both the protein and its background buffer. With 0 M urea concentration, there was a positive peak between 3500 cm^{-1} and 3000 cm^{-1}, corresponding with the N–H and O–H stretching vibrations within the protein backbone, an indicator of stable protein (Figure 4.25D). The presence of low intensity peaks between 1750 cm^{-1} and 1500 cm^{-1} range corresponds with the amide I and II bands respectively, which are indicative of the protein's secondary structural elements such as alpha-helices and beta-sheets. These peaks may indicate alterations or reductions in the content of these ordered structures within the protein. This was also supported by the negative peaks observed between 1250 cm^{-1}, and 1500 cm^{-1}, corresponding to the amide III region, potentially due to protein aggregation or partial unfolding (Tintor *et al.* 2024). A distinct peak at 3700 cm^{-1} further supported disruption of the protein's native hydrogen-bonding network.

At 4 M and 5 M urea concentration (Figure 4.25D), the FTIR spectra showed similar overall profile, with different in specific peak intensity and position, indicating progressive secondary structure. The spectrum at 4 M suggested a more preserved secondary structure compared to 5 M, which shows greater disruption. This trend was consistent with the Far-UV circular dichroism data, where 4 M retains more structure while 5 M showed partial loss. An absorbance peak at 1600 cm^{-1} was observed a representation of Amide I band, suggesting the presence of secondary structure content such as alpha helices, beta sheets, and random coil which is consistent with Krimm and Bandekar, (1986), who stated that Amide I band (1700 cm^{-1}–1600 cm^{-1}) is the most sensitive identified of secondary structure and has noted its sensitivity to C=O stretching vibrations. Additionally, the absorbance peak observed at 1545 cm^{-1}, represented the Amide II band, showing a partial loss of protein backbone structure, suggesting the structural difference to 0 M urea concentration. This aligned with the findings by Murphy *et al.* (2014) who highlighted the role of the Amide II band (~1545 cm^{-1}) in ATR-FTIR spectroscopy, as well as its relevance in far-UV circular dichroism spectroscopy. The Amide II band, which arises primarily from N–H bending and C–N stretching vibrations, served as an indicator of protein backbone integrity.

Changes in these bands represented structural alterations, particularly during protein unfolding or denaturation. The negative peak at 3750 cm^{-1}, showed the disruption of the hydrogen bonds which arise from the O–H stretching region, an indicator of water–protein interactions, and the negative shift was due to the presence of urea or protein denaturation. At 7 M urea, the spectrum showed predominant negative peaks across all regions, indicating substantial loss of structured elements. The amide I and II bands were significantly reduced, consistent with widespread unfolding and the loss of alpha-helices and beta-sheets. This finding confirmed that high urea concentrations result in complete structural disruption, aligning with the Far-UV circular dichroism spectrum data.

Together, the ATR-FTIR and Far-UV circular dichroism results provide complementary evidence of VP6 wild-type protein's progressive destabilization under chemical denaturation. Mild urea concentrations (4 M) temporarily stabilize the secondary structure by preventing aggregation, while higher concentrations (5–7 M) result in stepwise loss of structural integrity. These findings provide critical insight into the folding landscape and denaturation profile of VP6 protein and support its inherent aggregation tendency in the absence of stabilizing conditions.

4.8.2 VP6 Mutant Protein Secondary Structure

To evaluate the structural stability of the VP6 mutant protein under denaturing conditions, Far-UV circular dichroism (CD) and ATR–FTIR spectroscopy were performed using a range of urea concentrations (0 M, 4 M, 5 M, and 7 M) (Figure 4.26). These analyses provided insight into the impact of the introduced mutations on the protein's secondary structure and overall folding behaviour. The Far-UV circular dichroism spectrum of the VP6 mutant protein at 0 M urea revealed key spectral differences compared to the VP6 wild-type protein (Figure 4.26A). The 208 nm data, known as the most sensitive to alpha-helical and some beta-structures, was only analysed for the 0 M urea-treated protein as the spectra at 4 M, 5 M, and 7 M were unreliable due to noise below 210 nm. At 0 M urea concentration a distinct trough at 218 nm was observed, indicator of beta-sheet content, while the absence of a distinct alpha-helical signal at 222 nm, suggested a reduction in alpha-helical structure.

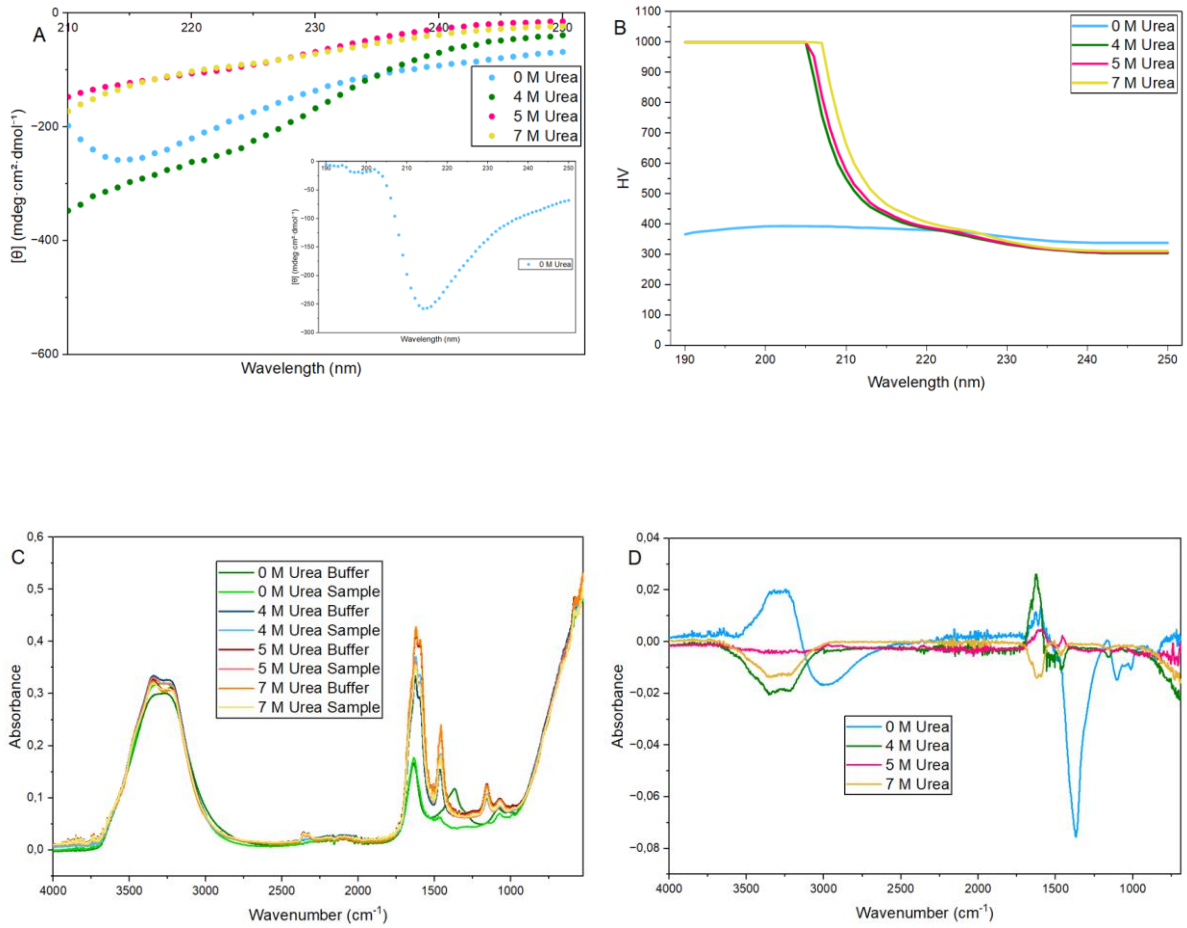


Figure 4.26: VP6 Mutant Protein Secondary Structure Characterisation

The VP6 mutant protein Far-UV circular dichroism spectra **(A)** were measured between the wavelengths of 250 nm to 190 nm. Then, protein was measured in urea concentrations of 0 M highlighted in blue, 4 M highlighted in green, 5 M highlighted in pink, and 7 M highlighted in yellow, and the HV using 6 µM protein **(B)**. The ATR–FTIR absorbance spectrum between the wavenumber of 4000 cm^{-1} to 500 cm^{-1} of the VP6 wild-type protein analysed using similar urea concentration from 0 M,4 M, 5 M, and 7 M. **(C)** shows overall measurement of the samples and buffer. **(D)** Shows the spectrum highlights of the protein-specific absorbance peaks more clearly.

Additionally, a noticeable shift in ellipticity towards 215 nm likely reflected the presence of protein aggregation or intrinsic misfolding. Unlike the VP6 wild-type protein, which displayed a broader signal near 205 nm, the VP6 mutant protein exhibited a more defined beta-sheet profile, indicating a possible alteration in aggregation behaviour or a shift toward alternative folding intermediates. These findings suggest that the introduced mutations may have modified the protein's folding pathway, favouring beta-rich aggregated forms over alpha-helical intermediates. The aggregates were also confirmed as per section 3.11.1, which proved that visible aggregates begin to form as the urea concentration drops below 4 M.

At 4 M urea concentration, the Far-UV circular dichroism spectrum indicated that the VP6 mutant protein maintains a significant level of the secondary structure, with clear negative ellipticity at both 222 nm (an indicator of the alpha-helices, suggesting a partially folded and stable helical regions) and 218 nm representing the presence of the beta-sheet. This suggested the presence of co-existing alpha-helices and beta-sheets under mildly denaturing conditions. Due to noise present below 210 nm, data at 208 nm were excluded from analysis even though it was representing beta sheets with structural analysis relying on the more stable 222 nm and 218 nm readings. Nonetheless, the spectral features above 210 nm were sufficient to support the conclusion that 4 M urea stabilized the VP6 mutant protein in a partially folded state, similar to the behaviour observed in VP6 wild-type protein. At 5 M and 7 M urea concentration, the VP6 mutant protein exhibited a significant loss of structural content, suggesting the unfolding of the protein with respect to its native helical conformation. The ellipticity at 222 nm reduced significantly, indicating loss of alpha-helical regions, while the 218 nm signal also weakened, suggesting unfolding or reorganization of beta-sheet elements. At 7 M urea concentration, the spectrum confirmed near-complete unfolding, consistent with denaturation-induced loss of ordered secondary structure. Compared to the VP6 wild-type protein, which retained partial structure at 5 M urea, The VP6 mutant protein exhibited a more rapid unfolding profile, suggesting a reduction in conformational stability.

High voltage (HV) values were analysed under all urea conditions. At 0 M urea concentration HV values between 250 nm and 190 nm wavelength range remained below 600 V, making the data acceptable and reliable for spectral analysis. At

wavelength between 210 nm and 190 nm for 4 M to 7 M urea concentration, HV reading exceeded 600 V, making the data unreliable due to excessive noise. However, from 250 nm to 210 nm, HV values remained below 600 V, signifying strong, reliable signals suitable for accurate structural interpretation, particularly at 222 nm and 218 nm. Overall, the VP6 mutant protein maintained partial secondary structure under mild denaturing conditions; however, higher concentrations of urea led to unfolding and loss of native conformation. Notably, the VP6 mutant protein begins to unfold at 5 M urea, whereas the VP6 wild-type protein requires 7 M urea, suggesting reduced stability and possible structural differences between the proteins. The difference in stability reflects structural changes which is based on the six amino acid substitutions differentiating the VP6 mutant protein from the VP6 wild-type protein. This suggested that the introduced mutations may have altered the protein's folding energy landscape, increasing sensitivity to denaturation.

ATR-FTIR spectroscopy analysis was used to further characterize the secondary structural content of the VP6 mutant protein (Figure 4.26C), showing the protein and their background buffer absorbance. Figure 4.26D showed different characteristically peaks that offered insight into the secondary structural state under denaturing conditions (0 M, 4 M, 5 M, and 7 M urea concentration). The analysis of the protein as the urea concentration increases showed a progressive loss of the secondary structure based on changes occurring in the Amide I (1600 cm^{-1}) and Amide II (1545 cm^{-1}) regions. There was a positive peak at 0 M urea concentration between 3500 cm^{-1} and 3000 cm^{-1} corresponding to N–H and O–H stretching vibrations, indicative of an intact hydrogen-bonding network and hydrated environment. The presence of the secondary structure was confirmed by the low intensity peak between 1750 cm^{-1} and 1500 cm^{-1}, a representation of the amide I and II bands respectively, which also represent the protein's secondary structural content such as alpha-helices and beta-sheets potentially altered or with reduced content of these organized structures in the proteins. This was supported by the negative peaks observed between 1250 cm^{-1} and 1500 cm^{-1}, corresponding to the amide III region, indicating structural destabilisation, likely arising from misfolding or aggregation. This was consistent with the Far-UV circular dichroism data showing aggregation-prone beta-rich structures at this condition.

At 4 M urea, positive peaks at 1600 cm^{-1} was observed, representing amide I band. The peak confirmed the presence of residual secondary structural content, consistent with a partially folded state, supported by a negative peak representing Amide II (1545 cm^{-1}) regions. Nonetheless, the appearance of several negative bands across the spectrum indicated early destabilization, suggesting that while the protein maintained some native features, the mutations conferred increased susceptibility to denaturation even at this intermediate urea concentration. At 5 M urea concentration, there was a less intense positive peak at 1600 cm^{-1} and 1545 cm^{-1}, showing the protein undergoing further destabilization becoming more unfolded. Compared to the VP6 wild-type protein, which maintained relatively stable peaks at this concentration, the VP6 mutant protein showed greater loss of structural integrity. At 7 M urea concentration, the spectrum was represented by the negative peaks across all key spectral regions, including Amide I, II, and III bands. This confirmed extensive denaturation and near-complete loss of ordered secondary structure, consistent with the Far-UV circular dichroism results. The spectral transition from partially folded to fully denatured forms occurred at lower urea concentrations for the VP6 mutant protein compared to the VP6 wild-type protein. The early onset of destabilization, evident at 4 M and 5 M urea, and complete structural loss at 7 M, supports the conclusion that the VP6 mutant protein is more sensitive to denaturing conditions.

The combined Far-UV circular dichroism and ATR–FTIR analysis revealed that both VP6 wild-type protein and VP6 mutant protein undergo progressive unfolding with increasing urea concentrations; however, their structural responses differed significantly. At 4 M urea, the wild-type protein exhibited relatively stable secondary structure with well-defined Far-UV circular dichroism and ATR-FTIR signatures, whereas the VP6 mutant protein showed early signs of destabilization. At 5 M urea, the VP6 wild-type protein maintained partial structural features, while the VP6 mutant protein showed considerable unfolding, indicated by reduced ellipticity and reduced ATR-FTIR band intensity. At 7 M urea, both proteins were fully denatured, with Far-UV circular dichroism and ATR-FTIR spectra confirming complete secondary structural loss.

4.9 Intrinsic Fluorescence

Intrinsic tryptophan fluorescence focuses on the conformational changes in the VP6 proteins, providing insights into their structural stability and folding behaviour under different denaturing conditions. The VP6 proteins were excited (λ_{ex}) at a wavelength of 295 nm, and their fluorescence emission spectra were recorded between the wavelengths of 500 nm to 280 nm at 22°C. As per Teale and Weber (1957), the tryptophan residues absorb UV at 295 nm and emit light at 350 nm free in solution or in proteins when fully exposed to solvent. Tryptophan (Trp) λ_{max} is highly influenced by its surrounding environment, varying from around 330 nm to 355 nm and generally reflects how exposed the chromophores are to the solvent (Vivian and Callis, 2001). The exposed tryptophan residues, typically associated with the VP6 protein unfolding, shows a peak emission maximum wavelength (λ_{max}) between 345 nm and 355 nm. In contrast, buried tryptophan residues found in a well-folded tertiary structure shows peak emission maxima from 330 nm to 340 nm. The VP6 wild-type protein and VP6 mutant protein contained five tryptophan residues (position 59, 139, 181, 247 and 367). The alpha helical domain of the protein contains three of the tryptophans (position 59,139 and 367) and the beta sheet domain contains two of the tryptophans (position 181 and 247) (Figure 1.4).

The structural stability of VP6 wild-type protein was further measured under increasing concentrations of urea (0 M, 4 M, 5 M, and 7 M) using intrinsic fluorescence spectroscopy with excitation at 295 nm (Figure 4.27). This approach probes the environment of tryptophan residues, which are sensitive indicators of tertiary structural changes or aggregation state. A high scattering signal at 295 nm under 0 M urea indicated the presence of large protein aggregates and significant turbidity under native conditions. In contrast, low signal intensities at 295 nm were observed in the presence of 4 M, 5 M, and 7 M urea, suggesting the absence of aggregation or scattering impurities and confirming a more soluble protein under these denaturing conditions.

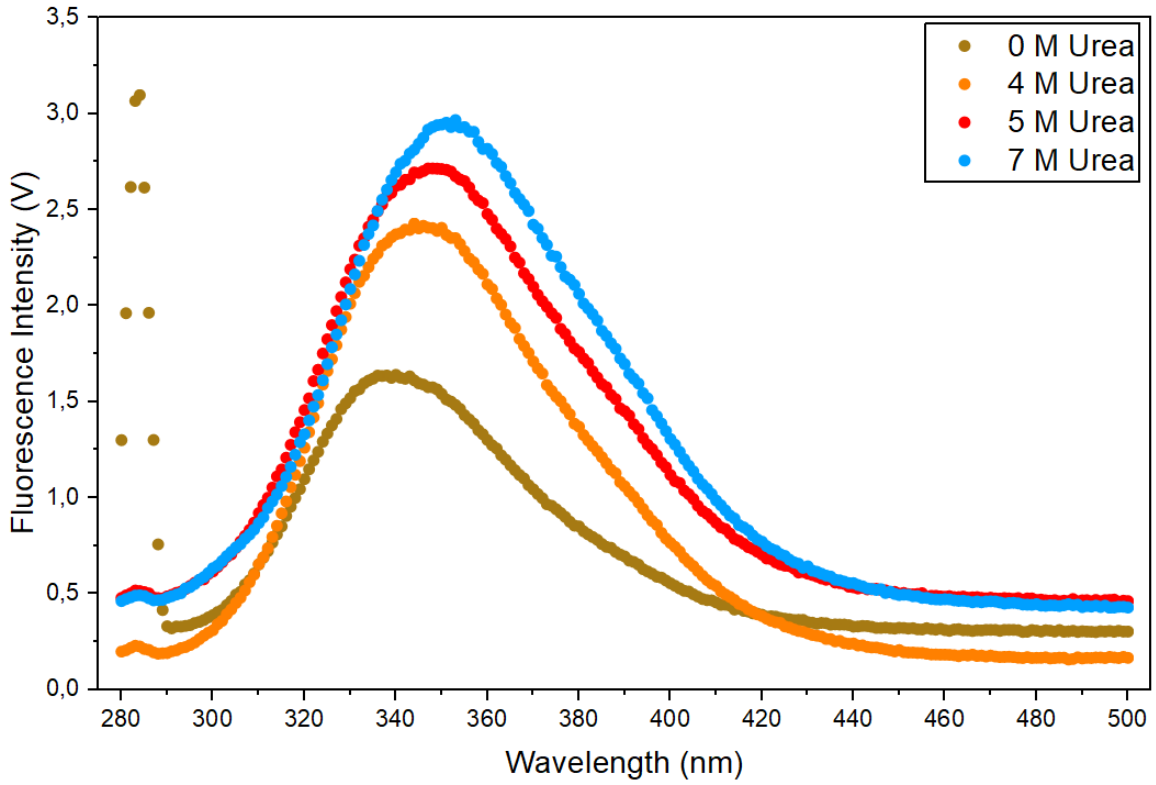


Figure 4.27: Intrinsic Fluorescence Spectra of the VP6 Wild-type Protein

The tryptophan residues emission spectra were measured using 8 µM protein and the protein was excited at 295 nm. Emission was measured between 500 nm and 280 nm to investigate changes in the protein's tertiary structure. Spectra were obtained under different urea concentrations: 0 M (brown), 4 M (orange), 5 M (red), and 7 M (blue).

Under native conditions (0 M urea), the fluorescence emission intensity at 340 nm was reduced, accompanied by high signal at the excitation wavelength (295 nm), indicative of strong light scattering from large aggregates. This suggested the existent of protein in an aggregated or multimeric state, likely due to its intrinsic tendency to self-assemble in the absence of denaturants, as also reported by Tao *et al.* (2023). The reduced emission intensity under native conditions reflects quenching effects caused by aggregation and the restricted solvent accessibility of tryptophan residues.

The maximum emission wavelength shift was observed as the urea concentration increased to 4 M, the emission maximum was 343 nm, suggesting partial unfolding and disruption of compact quaternary structure. This shift indicated previously buried tryptophan residues in hydrophobic core, becoming solvent-exposed, a characteristic of early unfolding transitions. Further increases in urea concentration resulted in continued red shifts: to 348 nm at 5 M and 354 nm at 7 M urea. This progressive change in emission maxima reflected increased solvent exposure of tryptophan residues and a gradual transition from folded oligomers to unfolded monomers. At 7 M urea, the near-complete red shift suggested that the protein has lost most of its tertiary structure, consistent with extensive unfolding confirmed by Moore *et al*. (2020), who observed progressive red shift in tryptophan fluorescence emission with 6 M, indicating increased solvent exposure of the chromophores.

In Figure 4.28, the emission spectra of the VP6 mutant protein showed consistency in the excitation maximum at 295 nm for 4M, 5 M, and 7 M of urea when compared to VP6 wild-type protein. With 0 M urea concentration showing high scattering signal at 295 indicating the presence of large protein aggregates whereas the spectra obtained under denaturing conditions showed low scattering at 295 nm, indicating no aggregation or scattering impurities. Under native conditions (0 M urea), a strong scattering signal was observed at 295 nm, indicating the presence of large aggregates, consistent with Rayleigh scattering principles which explains that scattering intensity is inversely proportional to the fourth power of the wavelength ($\propto \lambda^{-4}$). Therefore, shorter wavelengths scatter much more strongly than longer wavelengths (Lakowicz, 2006). This suggested that the VP6 mutant protein, like the VP6 wild-type protein, remains prone to aggregation under non-denaturing conditions. The emission intensity at 340 nm was also reduced, further supporting the presence of light-scattering aggregates.

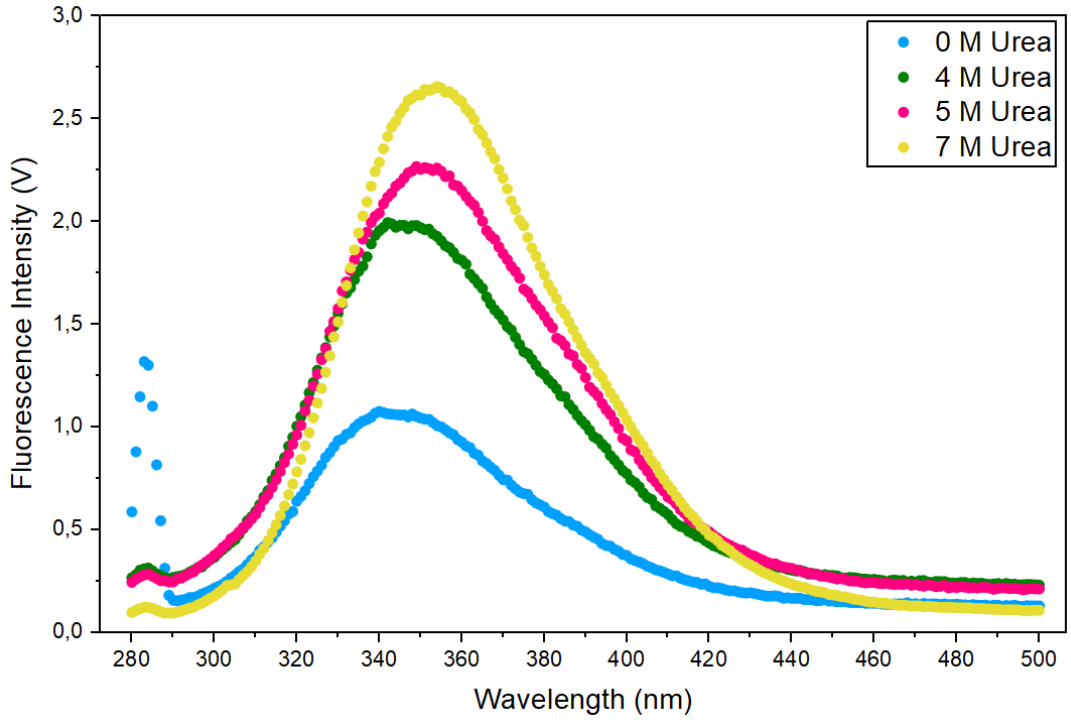


Figure 4.28: Intrinsic Fluorescence Spectra of the VP6 Mutant Protein

The tryptophan residues emission spectra were measured using 8 µM and the protein was excited at 295 nm. Emission was measured between 500 nm and 280 nm to analyse changes in the proteins tertiary structure. Spectra were obtained under different urea concentration: 0 M (blue), 4 M (green), 5 M (pink), and 7 M (yellow).

At 4 M urea, the emission peak was observed at 344 nm (highlighted in green) representing a partially folded protein relative to the native conformation, which typically exhibits emission maxima between 330 nm and 340 nm for tryptophan residues in a nonpolar environment. Unlike the wild-type protein, the VP6 mutant showed a more progressive and controlled unfolding behaviour under increasing denaturant concentrations. As urea concentration increased further, the emission maximum shifted to 349 nm at 5 M urea and to 355 nm at 7 M urea, suggesting full unfolding and a transition to a more extended, monomeric state.

Fluorescence data for both the VP6 wild-type protein (Figure 4.27) and VP6 mutant protein (Figure 4.28) exhibited similar behaviour. At 0 M urea, both showed significant aggregation and turbidity, indicated by high light scattering under native conditions. Increasing urea concentrations led to a red shift, showing unfolding and greater solvent exposure of tryptophan residues. The VP6 mutant protein exhibited this shift at lower urea concentrations, suggesting reduced structural stability. Although both proteins shared similar unfolding profiles, the VP6 mutant protein appeared more unfolded at 5 M and 7 M urea. Overall, intrinsic tryptophan fluorescence effectively monitored conformational changes and folding states under denaturing conditions.

4.10 Thermal Unfolding of the VP6 Protein

The thermal unfolding of both VP6 wild-type protein and VP6 mutant protein in 4 M urea concentration were monitored using Far UV circular dichroism and intrinsic tryptophan fluorescence spectroscopy. The temperatures were gradually increased from 20 °C to 94 °C, after which the same sample was cooled back to 20 °C. Additional measurements were taken following 24 hours incubation of the same sample at 4 °C to assess refolding or stability over time. For both, fluorescence and far-UV circular dichroism, spectra were recorded every one degree between 20 °C to 94 °C, then one measurement after cooling back same sample to 20 °C. A few temperatures between 20 °C to 94 °C were selected and plotted as the best representation of the VP6 protein behaviour during the thermal stress , with additional measurements taken after cooling the same sample back to 20 °C, followed by 24 hours incubation at 4 °C. The far-UV circular dichroism spectra were recorded in the range of 250 nm to 190 nm, although the most informative data were observed between 250 nm and 215 nm. The fluorescence spectrum was recorded between 500 nm and 280 nm using an excitation wavelength at 295 nm. Thermal unfolding and refolded proteins were monitored to

gain valuable insights into their structural stability and reversibility of folding after thermal denaturation. Aggregation observed at 0 M urea made it impossible for the protein to undergo thermal stress under native conditions. The use of urea reduces the protein's stability, causing it to unfold more easily even at lower temperatures (Zou *et al.* 1998).

4.10.1 Far-UV Circular Dichroism Thermal Unfolding

The thermal stability and secondary structure behaviour of the VP6 wild-type protein were investigated using far-UV circular dichroism (CD) spectroscopy in the presence of 4 M urea. Figure 4.29A showed the thermal denaturation profiles across a temperature range, while Figure 4.29B highlights representative spectra at selected temperatures. At 20°C, the negative ellipticity at both wavelengths (218 nm and 222 nm) indicated the presence of secondary structure despite the presence of 4 M urea. The signal at 222 nm represented the alpha-helical content and the signal at 218 nm represented alpha-helices with some beta-sheet and other ordered structural elements. This suggested that under mild denaturing conditions (4 M urea), the VP6 wild-type protein retains a partially folded conformation. There was a noticeable spectral transitional shift at 35°C, indicating the beginning of the secondary structure disruption. This temperature likely represented a midpoint in the unfolding pathway. At 40 °C, a near-complete loss of ellipticity at both 218 nm and 222 nm was observed, signifying extensive unfolding and transition towards a random coil conformation. Upon cooling the sample back to 20 °C, the ellipticity at both wavelengths (218 nm and 222 nm) did not recover, even after overnight incubation at 4 °C, indicating irreversible thermal denaturation.

High voltage (HV) measurements (Figure 4.29C) were used to assess the reliability of Far-UV circular dichroism data at selected temperatures. HV values exceeding 600 V indicated poor signal reliability due to increased sample turbidity, aggregation, or light scattering. At 20 °C, HV readings exceeded 600 V between 230 nm and 210 nm, indicating unreliable spectral data in this range. However, the region from 250 nm to 230 nm remained below the 600 V threshold, indicating acceptable data though less informative for secondary structure analysis.

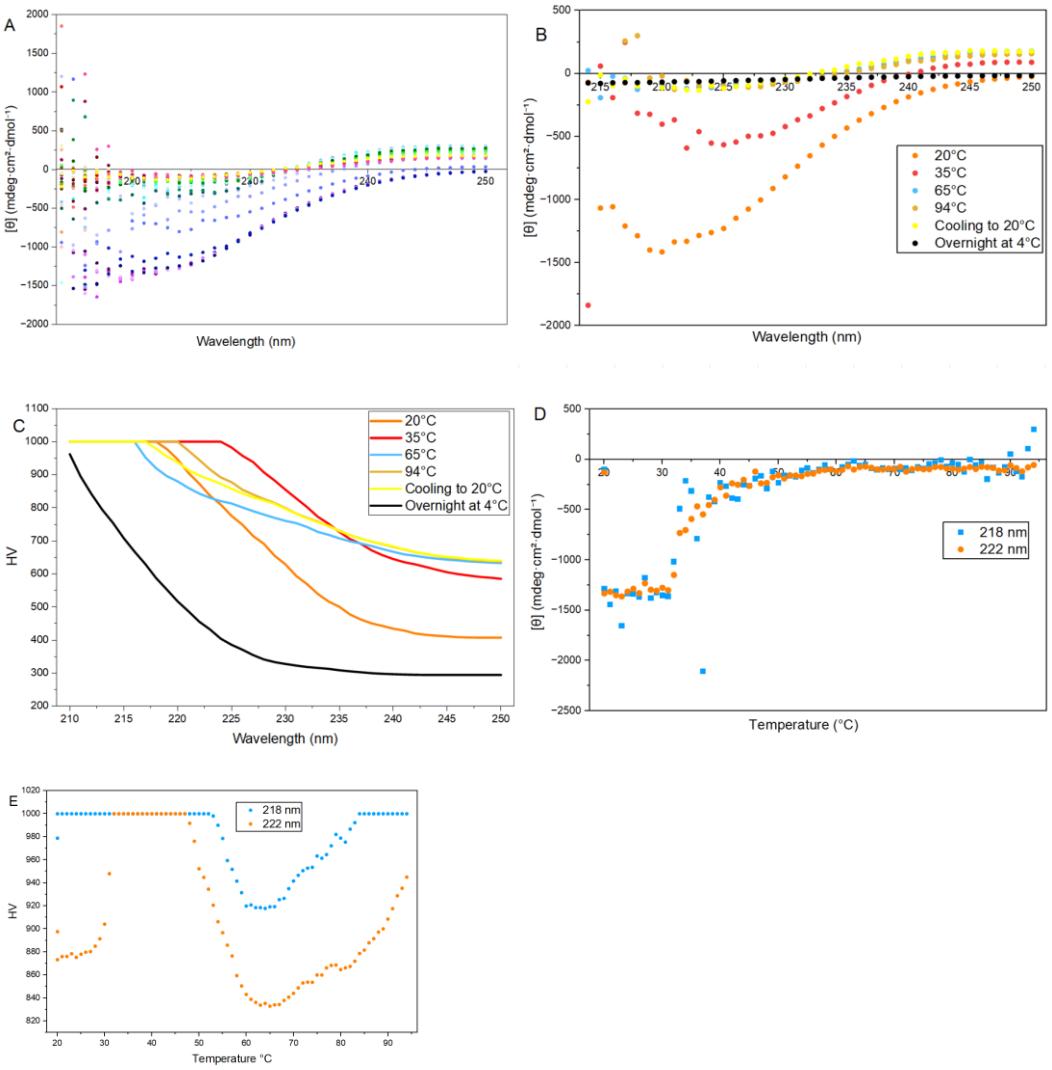


Figure 4.29: VP6 Wild-type Protein Far UV CD Thermal Unfolding

Spectra were measured between the wavelength of 250 nm to 215 nm using 6 µM protein in 4 M urea, to determine the secondary structure changes under thermal stress. **(A)** Spectra were measured at increasing temperatures from 20°C to 94°C and after cooling back to 20°C (20°C (—), 30°C (—), 40°C (—), 50°C (—), 60°C (—), 70°C (—), 80°C (—), 90°C (—), 94°C (—), and Cooling to 20°C (—)). **(B)** Showed selected spectra collected at different temperatures to illustrate changes in secondary structure. **(C)** Shows the accuracy and reliability of the Far-UV circular dichroism HV data. **(D)** Shows changes in ellipticity at 218 nm (blue) and 222 nm (orange), highlighting temperature-dependent unfolding and the structural stability of alpha-helices and beta-sheets regions. **(E)** Shows the accuracy, reliability, and aggregation through HV of the Far-UV circular dichroism spectra measured at wavelength of 218 nm highlighted in blue and 222 nm highlighted in orange.

After heating to 35°C and above, the HV readings remained above 600 V across the entire spectral range (250 nm to 210 nm), confirming persistent aggregation or structural loss. Even after cooling back to 20°C and overnight incubation at 4°C, HV readings remained above 600 V from 220 nm to 215 nm, suggesting incomplete clearance of large aggregates or sedimentation of insoluble material. Although HV values were lower in the 250–220 nm range, signal quality remained compromised and unreliable, confirming irreversible structural loss. This observation was consistent with the fact that thermal unfolding and irreversible denaturation interfered with the quality of the Far-UV circular dichroism signal by increasing light scattering. Moreover, high HV readings following thermal treatment at 35 °C and higher temperatures consistently reached or exceeded 600 V across all measured wavelengths, indicating significant protein aggregation. These findings supported the interpretation that thermal stress overrides the protective effects of 4 M urea, resulting in aggregation despite the presence of the denaturant.

Due to a poor signal-to-noise ratio confirmed by HV data, the 208 nm CD signal could not be used. While the 218 nm and 222 nm signals were also not fully reliable for assessing secondary structure, they were plotted to explore potential aggregation-related changes. The Far-UV circular dichroism spectra in Figure 4.29D supported this interpretation, showing a partially maintained secondary structure between 20 °C and 35 °C, followed by a sharp transition (gradual decline in ellipticity) beyond 35 °C. This cooperative decrease is consistent with the unfolding of residual secondary structure. Importantly, no recovery in spectral features were observed post-cooling, supporting the conclusion that the protein undergoes irreversible denaturation under thermal stress, even in the presence of 4 M urea.

Figure 4.29E further demonstrated the irreversible nature of unfolding, with a distorted or flattened Far-UV circular dichroism spectrum observed after the thermal cycle, characteristic of aggregated protein lacking ordered structure. A notable signal drop between 50°C and 80°C likely reflected the precipitation of large aggregates, as proteins are expelled from the optical path, a well-documented phenomenon in Far-UV circular dichroism thermal denaturation studies (Benjwal *et al.* 2006). These findings highlight the intrinsic thermal instability of the VP6 wild-type protein and its propensity to form irreversible aggregates under stress conditions.

Thermal denaturation of the VP6 mutant protein in 4 M urea was evaluated using far-UV circular dichroism spectroscopy. Figure 4.30A presents temperature-dependent far-UV circular dichroism spectra, while Figure 4.30B shows selected spectra at key temperatures of the thermal unfolding pathway (20 °C, 60 °C, 94 °C, after cooling to 20 °C, and overnight at 4 °C). At 20 °C, the VP6 mutant protein showed negative ellipticity at 218 nm and 222 nm, consistent with a partially folded conformation, confirming the presence of ordered secondary content such as beta-sheets (218 nm) and alpha-helices (near 220–222 nm) structural components. This indicates the protein remains in an ordered, though non-native state, in the presence of 4 M urea. There was a noticeable transitional shift after 46°C, showing the onset of cooperative secondary structure loss. This transition served as a critical midpoint of thermal stress, beyond which a complete loss of alpha-helical and beta-sheet content was observed. Spectra collected at 65°C and 94°C, showed a substantial loss of signal at both 218 nm and 222 nm, indicating a transition to a fully unfolded, likely random coil conformation. No recovery in ellipticity was observed after cooling the sample back to 20 °C or after overnight incubation at 4 °C, indicating irreversible thermal denaturation. Compared to the VP6 wild-type protein, which began unfolding at 35 °C, the VP6 mutant exhibited improved thermal resistance.

To assess signal reliability, HV data were collected across the spectral range (250–210 nm), as shown in Figure 4.30C. At 20 °C, HV values exceeded 600 V between 217 nm and 210 nm, indicating an unreliable signal in that region. However, the key analytical wavelengths at 218 nm and 222 nm remained below the 600 V threshold, confirming the reliability of these data points for structural interpretation. Notably, in contrast to the VP6 wild-type protein, the VP6 mutant displayed lower HV values across a broader temperature range, indicating improved spectral integrity and reduced light scattering due to lower aggregation levels during early unfolding. Even after thermal exposure to 45 °C and cooling back to 20°C, HV values at 222 nm remained below 600 V, further confirming the data's reliability. These observations suggested that the unfolding behaviour of the VP6 mutant protein reflects true structural transitions rather than artefacts introduced by aggregation or turbidity.

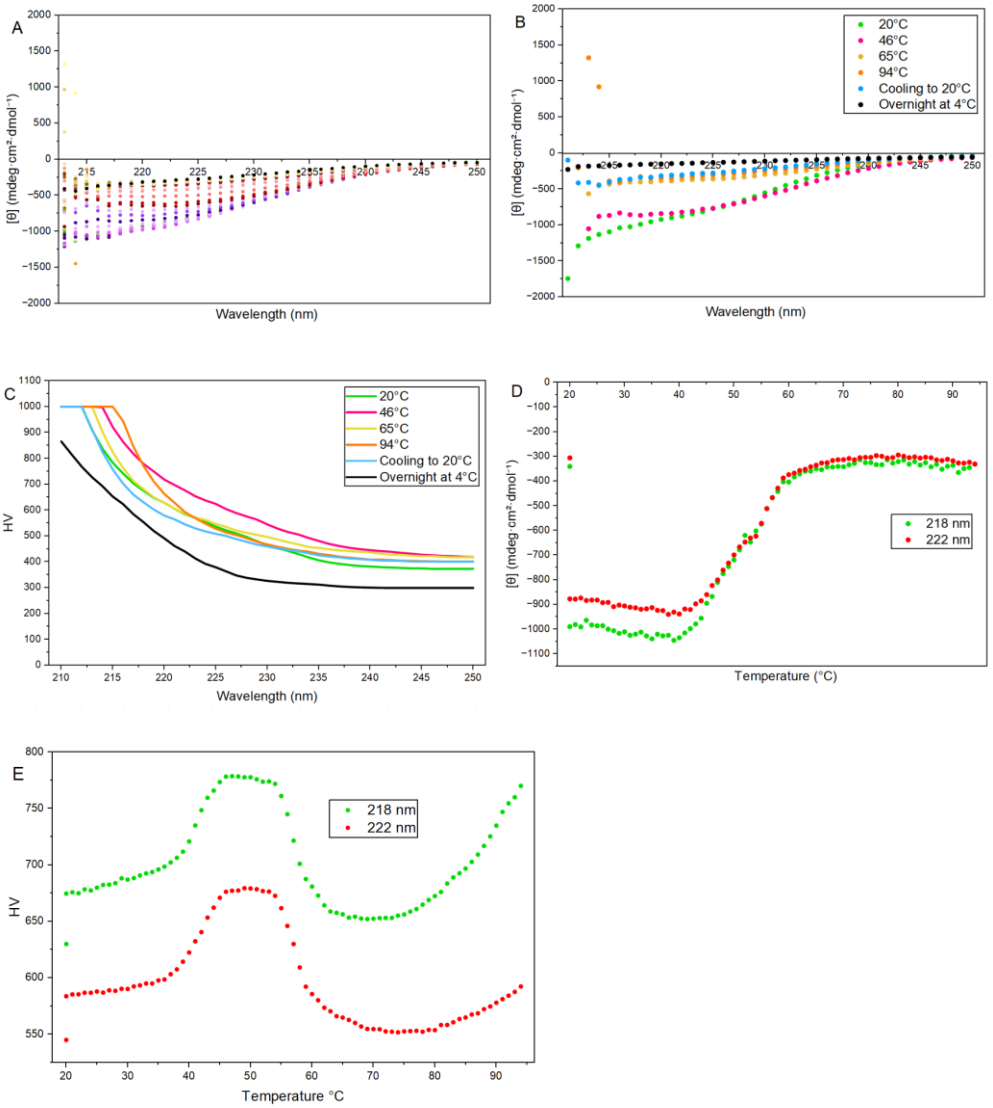


Figure 4.30: VP6 Mutant Protein Far UV Thermal Unfolding

The spectra were measured between 250 nm to 215 nm using 6 µM protein in the presence of 4 M urea concentration to determine the secondary structural changes during thermal stress. **(A)**. The spectra were collected as the temperatures increases from 20°C to 94°C and after cooling to 20°C (20°C (▬), 30°C (▬), 40°C (▬), 50°C (▬), 60°C (▬), 70°C (▬), 80°C (▬), 90°C (▬), 94°C (▬), and Cooling to 20°C (▬)). **(B)** Displays selected spectra collected at different temperatures to illustrate changes in secondary structure. **(C)** Shows the accuracy and reliability of the Far-UV circular dichroism HV data. **(D)** Highlights temperature-dependent unfolding, with focuses on the alpha-helical and beta-sheet regions at 218 nm (green) and 222 nm (Red). **(E)** Shows the accuracy, reliability and aggregation through HV of the Far-UV circular dichroism spectra data between wavelength of 218 nm highlighted in green and 222 nm highlighted in red.

The delayed onset of high HV values also supports the interpretation that the VP6 mutant protein was less prone to early aggregation compared to the VP6 wild-type protein. These differences were likely due to the six amino acid substitutions between the two proteins, which enhanced solubility or reduced aggregation propensity under thermal stress.

Figure 4.30D presented the ellipticity trends at 218 nm and 222 nm during the thermal cycle. From 20°C to 45°C, the ellipticity values remained relatively stable, consistent with a stable secondary structure. Between 45°C and 55°C there was a visible change in protein unfolding, followed by a sharp decline in ellipticity up to 94°C, consistent with complete loss of ordered structure. The absence of signal recovery upon cooling was consistent with irreversible denaturation and aggregation. The gradual unfolding profile, without any indication of reversible intermediates, further supports the conclusion that the VP6 mutant protein does not refold under the tested conditions but instead undergoes irreversible aggregation. However, compared to the VP6 wild-type protein, the VP6 mutant protein showed improved thermal resistance, requiring higher temperatures to reach this point, indicating improved thermal stability and reduced aggregation propensity. This suggested that the introduced mutations may have altered intramolecular interactions, enhancing stability under thermal stress. However, this benefit did not extend to reversible folding, as the protein ultimately aggregated, its delayed response to thermal stress reflects a more stable folded state under denaturing conditions.

Figure 4.30E showed the ellipticity and HV trends following a familiar pattern. Between 20°C and 45°C, the ellipticity remained stable and HV values were within acceptable limits, confirming the presence of structured protein. After 45°C, the HV began increasing, indicating increased turbidity due to protein aggregation. Interestingly, a subsequent drop in HV around 60 °C was observed, which likely corresponds to the precipitation of large aggregates out of solution, thereby reducing light scattering within the optical path. Thus, the combined HV and ellipticity behaviour confirmed that the observed changes are not instrumental artifacts but correspondence of genuine aggregation and irreversible structural loss, with cooling failing to restore the folded state. Overall, the VP6 mutant protein demonstrated enhanced thermal stability compared to the VP6 wild-type protein, unfolding at higher temperatures and maintaining spectral integrity longer. However, the unfolding remained irreversible,

and the mutant ultimately aggregated under thermal stress. These findings suggest that while the introduced mutations may reduce aggregation propensity and stabilize the folded state, they do not prevent misfolding or irreversible denaturation under elevated temperatures.

4.10.2 Fluorescence Monitoring of Rotavirus VP6 Thermal Unfolding

Intrinsic tryptophan fluorescence focuses on the tertiary environment surrounding the tryptophan residues in the VP6 proteins and the changes in their tertiary structure during the thermal unfolding and folding. The thermal unfolding and refolding behaviour of the VP6 wild-type protein in 4 M urea concentration was studied using intrinsic tryptophan fluorescence (Figure 4.31A) by monitoring the changes in tertiary structure and aggregation during temperature-induced stress. Spectra were recorded at selected temperatures 20°C, 35°C, 65°C, 94°C followed by cooling to 20°C and overnight at 4°C (Figure 4.31B). At 20°C, the emission maximum was observed at 344 nm, indicative of partially buried tryptophan residues within a hydrophobic environment, consistent with a folded native like conformation and low light scattering (Lakowicz, 2006). Upon heating to 35°C, a slight blue-shift to 342 nm was observed, suggesting maintenance of a hydrophobic environment around the tryptophan residues. However, a reduction in fluorescence intensity suggested possible early structural destabilization or the onset of aggregation, which can quench emission or increase light scattering (Vivian and Callis, 2001).

At 65°C, the emission maximum returned to 344 nm, accompanied by a sharp drop in fluorescence intensity, suggesting a destabilization of tertiary structure, with possible exposure of tryptophan's to solvent and/or formation of early aggregates that quench fluorescence or increase scattering. Notably, at 94°C, an unexpected blue shift to 339 nm was observed, which was unusual for thermal denaturation. Unlike typical red shifts seen in fully unfolded proteins, this blue shift suggested burial of tryptophan residues within newly formed hydrophobic aggregates, rather than exposure to solvent, a sign of non-native aggregation. The emission maximum upon cooling back the same sample to 20°C shifted to 348 nm with persistently low intensity. The red shift compared to the native state indicated incomplete refolding and more solvent exposure of tryptophan residues. The reduced fluorescence intensity suggested misfolded intermediates or residual aggregation, which quench fluorescence or scatter light.

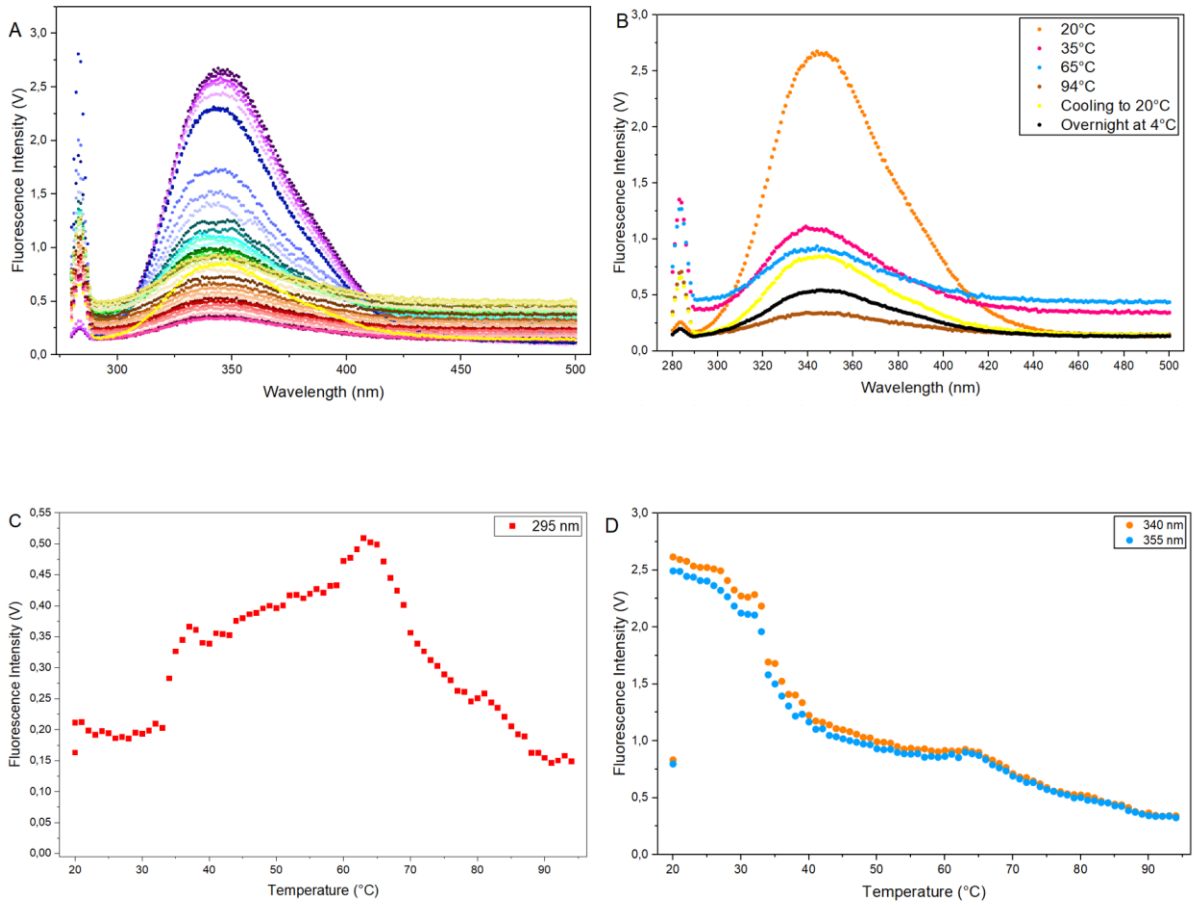


Figure 4.31: VP6 Wild-type Protein Thermal Unfolding

The measurement of the tertiary structure changes as monitored using the intrinsic fluorescence versus temperature during heat denaturation using 8 µM protein. **(A)** The fluorescence emission spectra were measured between 280 nm to 500 nm from 20°C to 94°C and cooling back to 20°C (20°C (—), 30 °C (—), 40 °C (—), 50 °C (—), 60 °C (—), 70 °C (—), 80 °C (—), 90 °C (—), 94 °C (—), and cooling to 20°C (—)). The fluorescence focused on the tryptophan which was excited at 295 nm. **(B)** The fluorescence emission spectra showing tryptophan spectra overview at selected temperature 20°C to 94°, cooling back to 20°C and overnight incubation at 4°C. **(C)** Shows thermal fluorescence analysis monitoring tryptophan emission following excitation at 295 nm. **(D)** The fluorescence emission intensity at 340 nm and 355 nm.

After overnight incubation at 4°C, the emission maximum remained red-shifted at 346 nm, with reduced fluorescence intensity, indicating burial of tryptophan residues, likely within non-native, aggregated structures. While this suggests altered tertiary environment, it does not confirm irreversibility on its own. However, Far-UV circular dichroism (Figure 4.29) ellipticity at both 218 nm and 222 nm did not recover, confirming irreversible thermal denaturation of the secondary structure.

Complementary fluorescence measurements using 295 nm excitation (Figure 4.31C) showed two-phase unfolding process. Between 20°C and 30°C, fluorescence emission initially decreased gradually between 20°C and 30°C, possibly indicating early structural rearrangements within the protein, losing the tertiary contacts or local unfolding. This was followed by a further decline in fluorescence intensity from 30°C to 60°C, consistent with the onset of aggregation, as indicated by the increased light scattering. These two distinct transitions (initial rearrangement followed by aggregation) suggested a multi-step unfolding process for the protein. These interpretations were consistent with Far-UV circular dichroism HV data, which revealed two separate thermal transitions corresponding to the changes in secondary structure (loss of secondary structure) and aggregation. The peak observed at 60°C may not solely reflect maximum tryptophan exposure; rather, it likely indicated increased light scattering at 295 nm resulting from the formation of aggregates or oligomers in solution, suggesting early stages of protein aggregation accompanying structural changes. Above 60°C, light scattering decreased likely due to protein precipitating out of solution. These resulted from fluorescence quenching associated with the burial of tryptophan residues within precipitated or aggregated structures. This data suggested that protein unfolding, and subsequent aggregation or precipitation are coupled processes, as the loss of secondary structure observed by Far-UV circular dichroism coincides with a drop in fluorescence intensity, followed by a decrease in light scattering possibly due to larger aggregates precipitating out of solution and no longer contributing to scattering or early aggregation limiting further structural transitions.

These findings suggested that the V6 wild-type protein transitioned early into aggregation-prone intermediates, which limited subsequent structural unfolding. The emission intensities at 340 nm (highlighted in orange) decreased consistently with increasing temperature, while no shift towards 355 nm (highlighted in blue) (Figure 4.31D). This indicated that, rather than fully unfolding, the protein likely adopted

intermediate or aggregated conformations that shielded tryptophan residues from solvent exposure throughout the thermal ramp. Overall, the results demonstrated that the VP6 wild-type protein experienced thermal stress in the presence of 4 M urea, leading to the formation of irreversible aggregates and structural rearrangements that prevented the recovery of its native conformation, potentially compromising its functional applications.

Thermal unfolding and refolding behaviour of the VP6 mutant protein in the presence of 4 M urea concentration was studied by monitoring the tertiary structural changes from 20°C to 94°C, followed by cooling back to 20°C (Figure 4.32A). Key temperatures selected were (20°C, 46°C ,60°C, 94°C, cooling to 20°C and overnight at 4°C) used to assess the protein's thermal stability and structural dynamics (Figure 4.32B). At the initial temperature of 20°C, the VP6 protein was expected to be in its native, folded conformation, with both secondary and tertiary structures intact. The emission maximum was observed at 344 nm, indicating a partially folded protein. Typically, native tryptophan emission maxima lie between 330 nm and 340 nm in a nonpolar environment suggesting some degree of solvent exposure and incomplete folding even at baseline conditions. Upon heating to 46°C, a blue shift to 339 nm was observed, associated with structural rearrangement, where tryptophan residues become momentarily more buried within hydrophobic pockets. This could reflect early unfolding events or, alternatively, quaternary structural changes due to aggregation, where burial of tryptophan residues occurred without global unfolding. At 65°C, the fluorescence profile suggested either complete structural rearrangement or a fully aggregated protein. Interestingly, at the maximum temperature of 94°C, the emission remained at 343 nm. Instead of the expected red shift to 345 nm to 355 nm characteristic of complete denaturation, this persistent intermediate wavelength indicated that tryptophans were still partially buried, likely due to aggregation shielding tryptophan residues by limiting solvent accessibility. Unlike the VP6 wild-type protein, which displayed a more cooperative, single transition at lower temperatures, the VP6 mutant protein exhibited at least two distinct transitions. This suggests a decoupling between unfolding and aggregation processes; however, the absence of a clear unfolding transition indicates that complete denaturation may not have been achieved, even at higher temperatures.

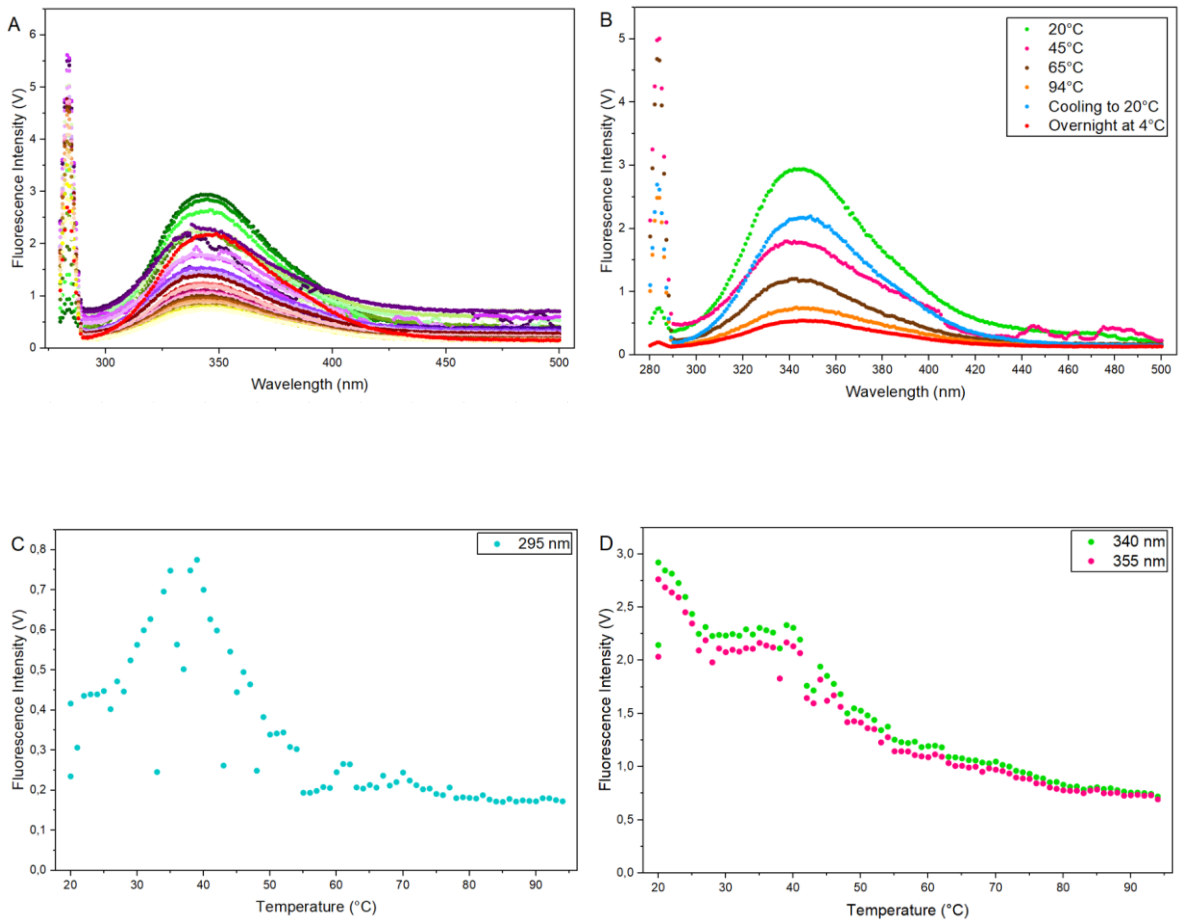


Figure 4.32: VP6 Mutant Protein Thermal Unfolding

The measurement of the tertiary structure changes using intrinsic fluorescence as temperature increases using 8 µM protein. **(A)** The fluorescence emission spectra were measured between 280 nm to 500 nm, increasing temperature from 20°C to 94°C and cooling back to 20°C (20°C (—), 30 °C (—), 40 °C (—), 50 °C (—), 60 °C (—), 70 °C (—), 80 °C (—), 90 °C (—), 94 °C (—), and cooling to 20°C (—)). The fluorescence focused on the tryptophan which was excited at 295 nm. **(B)** The fluorescence emission spectra showing tryptophan spectra overview at selected temperature 20°C to 94°, cooling back to 20°C and overnight incubation at 4°C. **(C)** Shows thermal fluorescence analysis monitoring tryptophan emission following excitation at 295 nm. **(D)** Emission intensity at 340 nm and 355 nm plotted across the temperature range to evaluate changes in the local environment of tryptophan residues.

Upon cooling back to 20°C, the emission maximum shifted to 349 nm, indicating a partial structural recovery. The tryptophan residues remained slightly more solvent-exposed than in their native state, resulting in an incomplete recovery of the protein's tertiary structure. Additionally, a decrease in fluorescence intensity suggested either persistent aggregation or solvent quenching of exposed fluorophores. This was also supported by the emission maximum at 348 nm after overnight incubation at 4°C, which showed the presence of non-native intermediates or misfolded aggregates that failed to refold to their conformational form. These findings suggest that the VP6 mutant protein formed stable, non-reversible aggregates during thermal stress.

For further understanding of the VP6 mutant protein tertiary structure transitions, fluorescence intensity vs temperature was plotted for wavelengths 295 nm, 340 nm, and 355 nm across the temperature range. At the 295 nm wavelength, an increase in fluorescence scatter was detected between 20°C and 50°C, consistent with the formation of larger aggregates that scatter more light, rather than indicative of tryptophan exposure (Figure 4.32C). As aggregation begins, these exposed tryptophan's may become buried within protein aggregates, resulting in enhanced light scattering and altered fluorescence signals. Interestingly, a decrease in fluorescence scatter after 50°C suggested a transition aggregation behaviour either due to dissociation of large aggregates into smaller, less light-scattering particles, or precipitation of insoluble species out of the optical path. This indicates a potential dissociating of structural unfolding from aggregation behaviour at higher temperatures. Notably, around 60°C, a decrease in scatter intensity reflected a transitional phase marked by structural reorganization or partial disaggregation a behaviour not observed in the VP6 wild-type protein. In contrast, the VP6 wild-type protein maintained higher fluorescence scatter throughout heating, indicating dominance of stable aggregates formation and limited conformational flexibility. These differences indicate that while VP6 mutant protein adopted a more dynamic, aggregation-prone intermediate more sensitive to thermal disruption, the VP6 wild-type protein aggregates were more heat-stable and resistant to further unfolding, supporting the idea that aggregation is a dominant and possibly irreversible during thermal denaturation.

Spectral monitoring at 340 nm (highlighted in green) and 355 nm (highlighted in pink) further supported this interpretation (Figure 4.32D). A progressive red shift in fluorescence emission maximum from 340 nm towards 355 nm, was consistent with

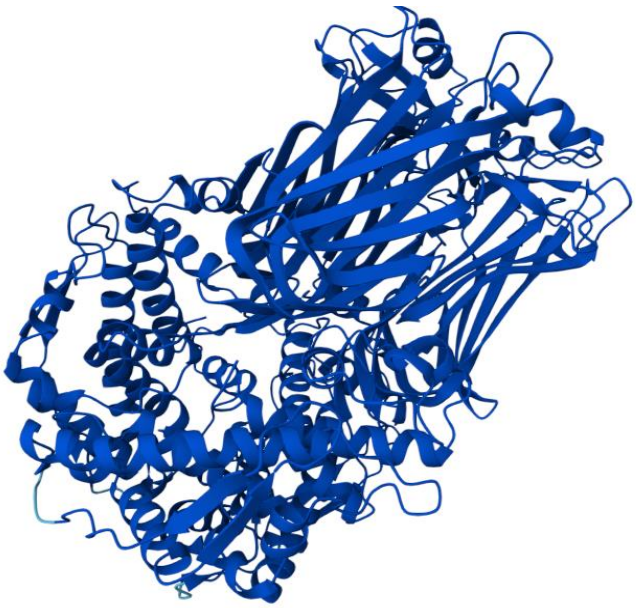
increased exposure of tryptophan residues to polar solvent environment during unfolding. The apparent decrease at 340 nm and corresponding increase at 355 nm reflects this shift in local environment rather than a simple fluorescence quenching. In addition to the spectral shift, increased temperature can also contribute to dynamic quenching by enhancing molecular motion and collision frequency, which weakens the excited state of tryptophan. These combined effects result in an unfolding-induced environmental change and temperature-dependent dynamic quenching, which results in an overall reduction in fluorescence intensity at specific wavelengths. The inability to restore fluorescence emission to its original state upon cooling indicated that denaturation was irreversible under the tested conditions. Aggregation, likely initiated during the heating phase, appears to have obstructed proper refolding, resulting in a non-native end state. Overall, thermal unfolding of the VP6 mutant protein in 4 M urea demonstrated incomplete denaturation, formation of misfolded intermediates, and an irreversible loss of native tertiary structure upon cooling. These findings underscore the complexity of VP6 folding dynamics and suggest that while mutations altered aggregation behaviour, they were insufficient to enable full structural recovery or solubility under thermal stress.

4.11 Quaternary Structure Characterization

The quaternary structure of the VP6 protein was evaluated using both *in silico* structural prediction and *in vitro* experimental validation to assess its oligomeric state and the potential impact of introduced mutations on its assembly behaviour.

The quaternary structure of the VP6 protein was predicted using AlphaFold 3 (DeepMind, 2024), generating a high-confidence model prediction with a pTM score of 0.9 and an ipTM score of 0.9 (Figure 4.34A). The pTM score reflects the overall fold accuracy of the protein complex, while the ipTM score assesses the relative positioning of the subunits. To explore potential higher-order oligomeric states, additional models were generated using 2, 4, 5, and 6 VP6 subunits (copies). The 2-subunit model produced a pTM score of 0.86 and an ipTM score of 0.82, indicating a reasonably high-confidence dimeric interaction, though slightly lower than the trimer model.

A



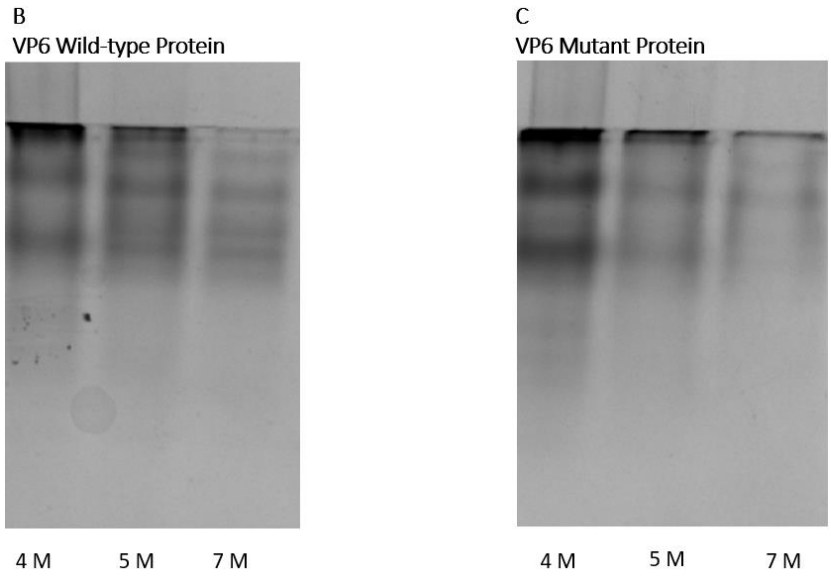


Figure 4.33: VP6 Protein Quaternary structure

AlphaFold 3 (DeepMind, 2024) predicted the quaternary structure of VP6 as a trimer, providing structural insights into its native oligomeric assembly **(A)**. The native PAGE of purified VP6 wild-type protein **(B)** and VP6 mutant protein **(C)** in different urea concentration (4 M, 5 M, and 7 M) was also performed to confirm the oligomeric state of the protein.

The 4-subunit model showed moderate confidence with a pTM score of 0.65 and an ipTM score of 0.7, suggesting a plausible but less stable assembly compared to the trimer. However, the models with 5 and 6 subunits produced both pTM and ipTM scores below 0.5, indicating low-confidence predictions and likely non-native or unstable assemblies. Given that the native form consists of two trimers, this suggests that AlphaFold3 did not accurately predict the physiologically relevant oligomeric state, instead modelling alternative, non-physiological assemblies. The VP6 mutant sequence produced similar structural predictions to the consensus (wild-type) sequence, despite containing six amino acid substitutions. Both the pTM and ipTM scores remained comparable across the modelled assemblies, suggesting that these mutations did not significantly disrupt the overall fold or subunit interactions of the protein. This indicates a degree of structural robustness in the VP6 quaternary architecture, even in the presence of sequence variation. These results support the trimer as the most structurally stable and biologically relevant quaternary form of VP6, consistent with its known role in forming functional capsid subunits. The reduced confidence in higher-order models suggests that such assemblies may not form stable, discrete structures under physiological conditions and may instead reflect transient or aggregation-prone interactions. Both scores were derived from the TM-score metric, where values above 0.8 suggested highly reliable structural models (Zhang and Skolnick, 2004; Xu and Zhang, 2010). Thus, the predicted structure of VP6 protein adopted a stable and well-defined oligomeric architecture, supporting its known ability to form trimers and higher-order assemblies essential for its biological function (Mathieu *et al.* 2001). To experimentally assess the oligomeric state of VP6 in its native environment and support the in-silico prediction, Native PAGE was performed.

Native-PAGE is a non-denaturing electrophoresis technique used for analysis of the protein native or folded state, allowing the study of their quaternary structures such as dimers, trimers, and higher-order complexes (Nowakowski *et al.* 2014). To experimentally validate the computational predictions and assess the quaternary structural dynamics of the VP6 protein under different denaturing conditions, Native-PAGE analysis of the VP6 wild-type protein (Figure 4.33B) and VP6 mutant protein (Figure 4.34C) was performed. The Native-PAGE revealed distinct banding patterns under 4 M, 5M, and 7 M urea conditions, indicating differences in their native oligomeric states.

For the VP6 wild-type protein, multiple bands were observed at 4 M, 5 M, and 7 M urea concentrations (Figure 4.33B), indicative of a heterogeneous population of oligomeric species. These likely include monomers, trimers, and potentially higher-order oligomers or capsid-like assemblies. The persistence of multiple species even under denaturing conditions reflects the protein's inherent structural flexibility and robust intermolecular interactions. Such behaviour is consistent with earlier studies that demonstrated the ability of VP6 to self-assemble into various quaternary forms depending on environmental conditions (Mathieu *et al.* 2001). In contrast, the VP6 mutant protein displayed fewer and less intense bands under the same urea concentrations (Figure 4.34C), particularly at 5 M and 7 M. This reduction in band number and intensity suggests a decrease in oligomeric complexity and/or weaker subunit interactions under denaturing conditions. The presence of fewer discrete oligomeric forms implies that the mutations introduced into the VP6 protein may have disrupted key residues involved in stabilizing the quaternary interface, thereby impairing the assembly of higher-order structures. The combination of AlphaFold 3 structural prediction, native-PAGE results, and prior observations of full-length VP6 protein behaviour suggests that VP6 possesses intrinsic structural flexibility, enabling it to assemble into different quaternary arrangements in different solution.

5 Discussion

In this study the main aim was to reduce the aggregation tendency of the VP6 protein by introducing six mutations targeting predicted aggregation-prone regions. VP6 is known to form inclusion bodies when overexpressed in *E. coli*, as reported by Zhao *et al.* (2011), a behaviour consistent with its role in forming viroplasmic inclusions and assembling into large quaternary structures such as trimers and higher-order oligomers. Although several studies have reported insoluble expression of VP6 protein in *E. coli* (Bredell *et al.* 2016; Kuri and Goswami, 2024), Kgokolo *et al.* 2017 successfully expressed rotavirus VP6 protein in *E. coli* using a reverse-translated consensus gene cloned into the pET-28a (+) vector. Expression trials were conducted in both BL21(DE3) and KRX *E. coli* strains. While VP6 expression in BL21(DE3) resulted in protein accumulating in inclusion bodies consistent with previous studies, the protein expression in KRX cells under optimized conditions yielded VP6 protein predominantly in the soluble fraction, sufficient for lab-scale structural work, but not ideal for high-yield production. This demonstrates that careful selection of host strain and induction conditions can significantly enhance the solubility of recombinant VP6 protein in *E. coli*, suggesting that rational, subtle mutations could improve soluble expression by reducing its tendency to aggregate.

Computational analysis using AGGRESCAN and supporting literature (Kuri and Goswami, 2024) who identified 24 aggregation-prone regions in VP6 protein with 21 of which were surface-exposed, potentially influencing the formation of the inclusion body in *E. coli* expression systems. The six mutations introduced in the current VP6 mutant protein (Figure 4.2) may not have exceeded this threshold since the expression did not fully shift toward the soluble fractions. This was supported by both VP6 wild-type protein and VP6 mutant protein, which remained largely in the insoluble fraction in all tested conditions with BL21(DE3) and NiCo21(DE3) strains, demonstrating limited solubility (Figure 4.8 to Figure 4.11). While AGGRESCAN provides good predictions, it is important to note that decreasing aggregation propensity often requires targeting a critical number or specific cluster of residues (Conchillo-Sole *et al.* 2007). In this study, the focus was on limiting mutations to naturally occurring amino acid substitutions to preserve biological relevance and maintain protein stability. It is possible that more drastic substitutions could have further reduced aggregation, but

they would have carried a higher risk of compromising the protein's structure or function.

Thus, the limited solubility suggests the six mutations were insufficient to reduce VP6's intrinsic aggregation tendency. While neither protein expressed predominantly in a soluble form, the VP6 mutant protein showed improved sensitivity to urea and pH, suggesting milder denaturing conditions were more effective for its solubilisation (Figure 4.13B) showing an altered behaviour during solubilization when compared to the VP6 wild-type protein, which required much harsher urea concentrations for solubility (Figure 4.13A). This aligns with the findings by Kuri and Goswami, (2024), who reported that rotavirus A VP6 forms classical inclusion bodies in *E. coli,* which are aggregates of misfolded protein rather than native-like forms. Their study showed that even with expression optimization in *E. coli* BL21(DE3), VP6 remained insoluble, requiring high urea concentrations for effective solubilization. All these indicate that although the critical aggregation threshold was not crossed, the mutations did partially alter the intermolecular interactions driving aggregation, supporting previous observations that targeted mutagenesis can fine-tune protein folding landscapes even without yielding soluble expression (Sormanni *et al.* 2015). Although purification predominantly yielded VP6 protein, mass spectrometry identified co-purifying *E. coli* proteins: EF-Tu and PheS in the VP6 wild-type protein, and YcfT and MalE in the VP6 mutant protein, indicating incomplete removal of non-specific contaminants. This suggested that denaturing conditions may expose hydrophobic or charged regions, promoting non-specific binding during IMAC (Bornhorst and Falke, 2000), and reflecting subtle differences in folding or aggregation behaviour between the VP6 wild-type protein and VP6 mutant protein.

VP6 protein naturally forms trimers as part of its functional structure (Mathieu *et al.* 2001) and differentiating between physiological oligomerization and non-specific aggregation is essential. Native PAGE (Figure 4.33) and absorbance data (Figure 4.22) suggested that the VP6 mutant protein (Figure 4.22B) in 4 M urea showed a more uniform quaternary structure with reduced high-molecular-weight aggregates compared to the VP6 wild-type protein, confirming a shift away from uncontrolled aggregation towards more ordered oligomeric forms. This was further supported by baseline light scattering at 340 nm (Figure 4.22A), which was higher for the VP6 wild-type protein, suggesting the presence of larger aggregates or oligomers, which aligns

with the hypothesis that the introduced mutations may have disrupted hydrophobic patches or altered surface electrostatics responsible for non-specific interactions without fully inhibiting native trimerization (Kuri and Goswami, 2024). However, despite this structural reorganization, these changes were insufficient to produce soluble expression, pointing out the complexity of VP6 folding, where both oligomeric interface stability and global folding energy landscape contribute to aggregation.

Thermal denaturation data from far-UV circular dichroism (Figure 4.29 and Figure 4.30) and intrinsic fluorescence (Figure 4.31 and Figure 4.32) revealed key differences between VP6 wild-type protein and VP6 mutant protein. Notably, the VP6 mutant protein began unfolding at a higher temperature (45°C) compared to the VP6 wild-type protein (35°C), indicating greater thermal resistance. However, neither protein could refold upon cooling, suggesting irreversible aggregation or misfolding post-denaturation. This was consistent with light scattering (Figure 4.25 to Figure 4.28) and buffer exchange data (described in section 3.11.1), which showed aggregation at 0 M urea. This behaviour differs with the trend observed in urea-induced denaturation (Intrinsic Fluorescence), where the VP6 mutant protein was more susceptible to unfolding at lower concentrations of urea (5 M) (Figures 4.27) compared to the VP6 wild-type protein (Figure 4.28). These behaviours highlighted the detailed effects of the mutations, which seemingly resulted in increased thermal resistance but decreased chemical stability. Such opposing trends may reflect alterations in hydrogen bonding or long-range electrostatic interactions that are more critical under chemical stress than thermal stress (Ghosh and Dill, 2010). Furthermore, the presence of significant structural changes at 0 M urea, including loss of ordered secondary structure, supports the concept that intrinsic instability contributes to aggregation even without external denaturation.

Far-UV circular dichroism and ATR–FTIR data (Figures 4.25 and Figure 4.26) consistently showed that the VP6 mutant protein was more sensitive to urea inducing secondary structure loss, with unfolding occurring earlier and more completely at 5 M and 7 M urea compared to VP6 wild-type protein which showed a fully unfolded protein at 7 M. Since hydrogen bonds are a major stabilising force for alpha-helices and beta-sheets, particularly in urea-induced denaturation, this suggests that the mutations may have compromised specific hydrogen bonding networks within the protein. Literature on urea-protein interactions supports that even minor alterations in sequence can

significantly shift the equilibrium between native and unfolded states (Auton and Bolen, 2005). This links to the earlier point regarding structural cooperativity, as such hydrogen bonding disruption may reduce the cooperative folding/unfolding transitions of the VP6 mutant protein. Notably, even with 0 M urea, both Far-UV circular dichroism and FTIR spectra also revealed early signs of structural instability in the mutant protein, indicating inherent weakness in its secondary structure even under native conditions.

Taken together, the differences in secondary and tertiary structure behaviour under urea stress suggested that the VP6 mutant protein showed reduced cooperativity in its unfolding transitions. This was evident in both Far-UV circular dichroism and fluorescence data, where the VP6 wild-type protein showed more sigmoidal (cooperative) unfolding curves, while the mutant displayed broader transitions (Figures 4.25 to Figure 4.28). Such reduced cooperativity could indicate a more dynamic, with partially folded intermediate stating a phenomenon often associated with reduced stability and higher aggregation susceptibility under certain conditions (Uversky *et al.* 2000). Moreover, the formation of more stable quaternary aggregates in VP6 wild-type protein under thermal stress, as seen by blue-shifted tryptophan emission at 94 °C (Figure 4.31), differs with the partial disaggregation seen in the VP6 mutant protein at 60°C. This suggests that while the mutant is more prone to unfolding under chemical denaturation, its aggregates are less stable and more dynamic, possibly due to weaker intermolecular interactions. These findings may inform future engineering of VP6, particularly where reversible or dynamic oligomerization is desirable for applications in nanotechnology or vaccine delivery.

6 Conclusion

This study demonstrates the inherent aggregation tendency of both VP6 wild-type protein and VP6 mutant protein, with expression predominantly occurring in inclusion bodies despite targeted mutagenesis and requiring strong denaturants for solubilization. The six rationally designed mutations, aimed at reducing aggregation by modifying surface-exposed, aggregation-prone regions, did not result in soluble expression under native conditions. However, the VP6 mutant protein exhibited altered biophysical behaviour, including improved solubilization under milder denaturing conditions, delayed thermal unfolding, and reduced high-molecular-weight aggregation compared to the VP6 wild-type protein. These findings suggest that while the mutations partially modified the aggregation landscape, they were insufficient to shift the expression toward soluble forms or restore native-like folding post-denaturation. Structural analyses revealed that both VP6 wild-type protein and VP6 mutant Protein lacked the capacity to refold after denaturation, indicating intrinsic folding instability. The VP6 mutant protein showed reduced unfolding cooperativity and greater sensitivity to chemical denaturants, suggesting compromised hydrogen bonding or long-range electrostatic interactions. All these findings highlight the inherent trade-offs between improving solubility, reducing aggregation, and maintaining structural integrity in recombinant VP6 expression. The limited improvements achieved through conservative mutations suggest a need for more targeted, structure-informed reengineering approaches. Future design should align with specific functional goals, such as enhancing trimer stability, promoting monomer solubility, or supporting higher-order assembly (preserving assembly competence for VLPs formation) for vaccine or nanotechnology applications.

7 Future Directions

Building on the predictive work of Kuri and Goswami, (2024), future VP6 engineering will benefit from deeper integration of sequence-based aggregation predictors such as Aggrescan3D and TANGO with molecular dynamics simulations to map the structural impacts of mutations in aggregation-prone regions more precisely. Additionally, rational mutagenesis focused on improving hydrogen bonding networks or disrupting hydrophobic patches without compromising oligomerisation could enhance expression and solubility. Furthermore, purifications could be optimized by fine-tuning refolding buffers, pH conditions, and detergent usage to manipulate the altered solubility profile of the VP6 protein.

8 References

Abid, N., Chillemi, G., and Salemi, M., 2019. Coding-Gene Coevolution Analysis of Rotavirus Proteins: A Bioinformatics and Statistical Approach. *Genes,* 11(1): 28.

Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A.J., Bambrick, J., Bodenstein, S.W., Evans, D.A., Hung, C.C., O'Neill, M., Reiman, D., Tunyasuvunakool, K., Wu, Z., Zemgulytė, A., Arvaniti, E., Beattie, C., Bertolli, O., Bridgland, A., Cherepanov, A., Congreve, M., Cowen-Rivers, A.I., Cowie, A., Figurnov, M., Fuchs, F.B., Gladman, H., Jain, R., Khan, Y.A., Low, C.M.R., Perlin, K., Potapenko, A., Savy, P., Singh, S., Stecula, A., Thillaisundaram, A., Tong, C., Yakneen, S., Zhong, E.D., Zielinski, M., Zídek, A., Bapst, V., Kohli, P., Jaderberg, M., Hassabis, D., and Jumper, J.M., 2024. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 630(8016): 493–500.

Afchangi, A., Jalilvand, S., Mohajel, N., Marashi, S.M., Shoja, Z., 2019. Rotavirus VP6 as a potential vaccine candidate. *Rev Med Virol,* 29: e2027.

Aiyegbo, M.S., Sapparapu, G., Spiller, B.W., Eli, I.M., Williams, D.R., Kim, R., Lee, D.E., Liu, T., Li, S., Woods Jr, V.L., Nannemann, D.P., Meiler, J., Stewart, P.L., Crowe Jr, J.E., 2013. Human rotavirus VP6-specific antibodies mediate intracellular neutralization by binding to a quaternary structure in the transcriptional pore. *PLoS ONE,* 8: e61101.

Arias C.F., and López S., 2021. Rotavirus cell entry: not so simple after all. *Current Opinion in Virology,* 48: 42–48.

Asowata, O.E., Ashiru, O.T., Mahomed, S., Sturm, A.W., and Moodley, P., 2018. Influence of vaccination status and clinical, seasonal and sociodemographic factors on rotavirus prevalence in KwaZulu-Natal, South Africa. *South. Afr. J. Infect. Di,* 33(5): 1-7.

Auton, M., Bolen, D.W., 2005. Predicting the energetics of osmolyte-induced protein folding/unfolding. *Proceedings of the National Academy of Sciences of the United States of America*, 102(42): 15065–15068.

Bar-Zeev, N., Jere, K.C., Bennett, A., Pollock, L., Tate, J.E., Nakagomi, O., Iturriza-Gomara, M., Costello, A., Mwansambo, C., Parashar, U.D., Heyderman, R.S., French, N., Cunliffe, N.A., and Vaccine Effectiveness and Disease Surveillance Programme, Malawi (VACSURV) Consortium., 2016. *Population impact and effectiveness of monovalent rotavirus vaccination in urban Malawian children 3 years after vaccine introduction: Ecological and case-control analyses*. *Clinical Infectious Diseases, 62*(2): S213–S219.

Chen, Y., and Barkley, M.D., 1998. Toward understanding tryptophan fluorescence in proteins. *Biochemistry*, 37(28): 9976–9982.

Bohnert, H.J., Smith, R.P., and Johnson, A.R., 1986. Spectral data comparability–acceptance criteria definition by statistical analysis. *Encyclopedia of Industrial Biotechnology: Bioprocess, Bioseparation, and Cell Technology*, 1(1): 1–9.

Bornhorst, J.A., and Falke, J.J., 2000. Purification of Proteins Using Polyhistidine Affinity Tags. *Methods Enzymol,* 326: 245–254.

Benjwal, S., Verma, S., Röhm, K.H., and Gursky, O., 2006. Monitoring protein aggregation during thermal unfolding in circular dichroism experiments. *Protein Science, 15*(3): 635–639.

Brack, P.L., Merkel, O., and Schroeder, R., 2024. A rapid method to monitor structural perturbations of high-concentrated therapeutic antibody solutions using Intrinsic Tryptophan Fluorescence Emission spectroscopy. *European Journal of Pharmaceutics and Biopharmaceutics*, 201: 114377.

Baroroh, U., Muscifa, Z. S., Destiarani, W., Rohmatullah, F. G., and Yusuf, M., 2023. Molecular interaction analysis and visualization of protein-ligand docking using Biovia Discovery Studio Visualizer. *Indonesian Journal of Computational Biology*, 2(1). 20-30.

Barth, A., 2007. Infrared spectroscopy of proteins, *Biochim. Biophys. Acta,* 1767: 1073–1101.

Barth, H.G., Boyes, B.E., and Jackson, C., 1998. Size exclusion chromatography and related separation techniques. *Anal Chem,* 70(12): 251R–278R.

Bolanos-Garcia, V.M., and Davies, O.R., 2006. Structural analysis and classification of native proteins from *E. coli* commonly co-purified by immobilised metal affinity chromatography. *Biochimica et Biophysica Acta (BBA) – General Subjects, 1760*(9): 1304–1313.

Bredell, H., Smith, J.J., Prins, W.A., Gorgens, J.F., and van Zy, W.H., 2016. Expression of rotavirus VP6 protein: a comparison amongst *Escherichia coli*, *Pichia pastoris and Hansenula polymorpha*, 16(2): fow001.

Bugli, F., Caprettini, V., Cacaci, M., Martini, C., Paroni Sterbini, F., Torelli, R., Della Longa, S., Papi, M., Palmieri, V., Giardina, B., Posteraro, B., Sanguinetti, M., and Arcovito, A., 2014. Synthesis and characterization of different immunogenic viral nanoconstructs from rotavirus VP6 inner capsid protein. *International Journal of Nanomedicine*, *9*: 2727–2739.

Burgess, R.R., 2009. Refolding solubilized inclusion body proteins. *Methods Enzymol,* 463: 259–82.

Burke. R.M., Tate, J.E., Kirkwood, C.D., Steele, A.D., and Parashar, U.D., 2019. Current and new rotavirus vaccines. *Current Opinion in Infectious Diseases,* 32: 435–444.

Carcamo-Calvo, R., Munoz, C., Buesa, J., Rodriguez-Diaz, J., and Gozalbo-Rovira, R., 2021. The Rotavirus Vaccine Landscape, an Update. Pathogens, 10(5): 520.

Carija, A., Navarro, S., de Groot, N.S., Puiggalí, C., and Ventura, S., 2019. Protein aggregation into insoluble deposits protects from oxidative stress. *Redox Biology*, 12: 699–711.

Charpilienne, A., Lepault, J., Rey, F., and Cohen, J., 2002. Identification of rotavirus VP6 residues located at the interface with VP2 that are essential for capsid assembly and transcriptase activity*. J. Virol*, 76: 7822–7831.

Choi, J.H., Keum, K.C., Lee, S.Y., 2006. Production of recombinant proteins by high cell density culture of *Escherichia coli*. *Chem. Eng. Sci*, 61(3): 876–885.

Choi, E., Kim, E., Oh, Y., Shin, K., Kim, H., and Kim, C., 2002. Expression of rota virus capsid proteins VP6 and VP7 in mammalian cells using Semliki Forest virus-based expression system. *J. Microbial. Biotechnol*, 12(3): 463-469.

Conchillo-Solé, O., de Groot, N.S., Avilés, F.X., Vendrell, J., Daura, X., and Ventura, S., 2007. AGGRESCAN: a server for the prediction and evaluation of "hot spots" of aggregation in polypeptides. BMC Bioinformatics, 8(65): 1-17.

Coskun, K.A., Yurekli, N., Abay, E.C., Tutar, M., Al, M., and Tutar, Y., 2022. Structure- and Design-Based Difficulties in Recombinant Protein Purification in Bacterial Expression. *Biochemistry*. IntechOpen.

Crawford, S.E., Criglar, J.M., Liu, Z., Broughman, J.R., and Estes, M.K., 2019. COPII Vesicle Transport Is Required for Rotavirus NSP4 Interaction with the Autophagy Protein LC3 II and Trafficking to Viroplasms. *Journal of Virology*, 94: e01341-e19.

Crawford, S.E., Ramani, S., Tate, J.E., Parashar, U.D., Svensson, L., Hagbom, M., Franco, M.A., Greenber, H.B., Greenberg, H.B., O'Ryan, M., Kang G., Desselberger, U., and Estes M.K., 2017. Rotavirus infection. *Nat. Rev. Dis. Prim,* 3: 17083.

Dadi, M., and Yasir, M., 2022. Spectroscopy and Spectrophotometry: Principles and Applications for Colorimetric and Related Other Analysis, 81-102.

da Silva Junior, H.C., de Mendonça, M.C.L., de Souza Pereira, M.C., da Rocha, N.A., de Azevedo, M.L.B., Leite, J.P.G., de Moraes, M.T.B., 2012. Comparison of two eukaryotic systems for the expression of VP6 protein of rotavirus species A: transient gene expression in HEK293-T cells and insect cell-baculovirus system. *Biotechnology Letters,* 34: 1623–1627.

Dennehy, P.H., 2008. Rotavirus Vaccines: An Overview. *Clinical Microbiology Reviews,* 21(1): 198 – 208.

Dennehy, P.H., 2015. Rotavirus infection: A disease of the past? *Infectious Disease Clinics of North America, 29*(4): 617–635.

Desselberger, U., 2014. Rotaviruses. *Virus Research*, 190: 75-96.

Donato, C.M., and Bines, J.E., 2021. Rotaviruses and Rotavirus Vaccines. *Pathogens*, 10(8): 959.

Dormitzer, P.R., Greenberg, H.B., and Harrison, S.C., 2001. Proteolysis of monomeric recombinant rotavirus VP4 yields an oligomeric VP5 core. *Journal of Virology*, 75(16): 7339–7350.

Dos Santos Rodrigues, F.H., Garcia Delgado, G., Da Costa, T.S., and Tasic, L., 2023. Applications of fluorescence spectroscopy in protein conformational changes and intermolecular contacts. *BBA Advances*, 3: 100091.

Estes, M.K., and Kapikian, A.Z., 2007. Rotaviruses and their replication. In D. M. Knipe and P. M. Howley (Eds.), *Fields Virology,* 5: 1917–1974.

Estes, M.K., Kang, G., Zeng, C.Q., Crawford, S.E., and Ciarlet, M., 2001. Pathogenesis of rotavirus gastroenteritis*. Novartis Found symp,* 238: 82-96.

Fink, A., 1998. Protein aggregation: folding aggregates, inclusion bodies and amyloid. *Fold and Design,* 3: R9–23.

Francis, D.M., and Page, R., 2010. Strategies to Optimize Protein Expression in E. coli. *Curr Protoc Protein Sci,* 61(1): 5241–52429.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., and Bairoch, A., 2005. Protein identification and analysis tools on the ExPASy server. In J. M. Walker (Ed.), *The proteomics protocols handbook*. Humana Press, 571-607.

Giaquinto, C., Van Damme, P., Muet, F., Gothefors, L., Maxwell, M., Todd, P., and da Dalt, L., 2007. Clinical Consequence of Rotavirus Acute Gastroenteritis in Europe. 2004-2005: the REVEAL study. *Journal of Infectious Diseases,* 195: S26-S35.

Gil-Garcia, M., Bañó-Polo, M., Varejão, N., Jamroz, M., Kuriata, A., Díaz-Caballero, M., Lascorz, J., Morel, B., Navarro, S., Reverter, D., Kmiecik, S., Ventura, S., 2018 Combining Structural Aggregation Propensity and Stability Predictions To Redesign Protein Solubility. *Mol Pharm,* 15(9): 3846–3859.

Ghisaidoobe, A.B.T., and Chung, S.J., 2014. Intrinsic tryptophan fluorescence in the detection and analysis of proteins: A focus on förster resonance energy transfer techniques. *International Journal of Molecular Sciences,* 15: 22518–22538.

Ghosh, K., Dill, K.A., 2010. Cellular proteomes have broad distributions of protein stability. *Biophysical Journal*, 99(12): 3996–4002.

Gooran, N., and Kopra, K., 2024. Fluorescence-Based Protein Stability Monitoring. A Review*. International Journal of Molecular Sciences***,** 25(3): 1764.

Greenberg, H.B., and Estes, M.K., 2009. Rotaviruses: from pathogenesis to vaccination. Gastroenterology, 136:1939–1951.

Greenberg, H., McAuliffe, V., Valdesuso, J., Wyatt, R., Flores, J., Kalica, A., Hoshino, Y., and Singh, N., 1983. Serological analysis of the subgroup protein of rotavirus using monoclonal antibodies. *Infect Immun*, 39: 91–99.

Goormaghtigh, E., Ruysschaert, J.M., and Raussens, V., 2006. Evaluation of the information content in infrared spectra for protein secondary structure determination, *Biophys. J*. 90(8): 2946–2957.

Hallowell, B.D., Tate, J., and Parashar, U., 2020. An overview of rotavirus vaccination programs in developing countries. Expert Rev Vaccines, 19(6): 529–37.

Han, J., Yang, J., Wang, Y., and Li, Y., 2019. The adequate amount of sodium chloride in Protein A wash buffer for effective host cell protein clearance. *Protein Expression and Purification, 158*: 59–64.

Hartman, R.H., Cohen, A.L., Antoni, S., Mwenda, J., Weldegebriel, G., Biey, J., Shaba, K., de Oliveira, L., Rey, G., Ortiz, C., Tereza, M., Fahmy, K., Ghoniem, A., Ashmony, H., Videbaek, D., Singh, S. Tondo, E., Sharifuzzaman, M., Liyanage, J., Batmunkh, N., Grabovac, V., Logronio, J., Serhan, F., and NakamurA, T., 2022. Risk Factors for Mortality Among Children Younger Than Age 5 Years with Severe Diarrhea in Low- and Middle-income Countries: Findings from the World Health Organization-coordinated Global Rotavirus and Pediatric Diarrhea Surveillance Networks. *Clin infect dis,* 76(3): e1047–e1053.

He, F., 2011. Bradford Protein Assay. *Bio-Protocol,* 1(6): e45.

Heinimaki, S., Tamminen, K., Hytonen, V.P., Malm, M., Blazevic, V., 2020. Rotavirus Inner Capsid VP6 Acts as an Adjuvant in Formulations with Particulate Antigens Only. *Vaccines (Basel),* 8(3): 365.

Hixon, J., and Reshetnyak, Y.K., 2009. Algorithm for the Analysis of Tryptophan Fluorescence Spectra and Their Correlation with Protein Structural Parameters. *Algorithms*, 2(3): 1155–1176.

Hoshino, Y., Sereno, M.M., Midthun, K., Flores, J., Kapikian, A.Z., and Chanock, R.M., 1985. Independent segregation of two antigenic specificities (VP3 and VP7) involved in neutralization of rotavirus infectivity. *Proc Natl Acad Sci USA,* 82: 8701-8704.

Hu, L., Crawford, S.E., Hyser, J.M., Estes, M.K., Prasad, B.V., 2012. Rotavirus non-structural proteins: structure and function. *Curr Opin Virol*, 2(4): 380-388.

Hu, D., Qin, Z., Xue, B., Fink, A.L., and Uversky, V.N., 2008. The effect of methionine oxidation on structural properties, conformational stability, and aggregation of immunoglobulin light chain LEN. *Biochemistry, 47*(33): 8665–8677.

Imai, K., and Mitaku, S., 2005. Mechanisms of secondary structure breakers in soluble proteins. *Biophysics*, 1, 55–65.

Jere, K.C., O'Neill, H.G., Potgieter, A.C., and van Dijk, A.A., 2014. Chimaeric Virus-Like Particles Derived from Consensus Genome Sequences of Human Rotavirus Strains Co-Circulating in Africa. *PLOS One*, 9(9): e105167.

Johne, R., Schilling-Loeffler, K., Ulrich, R.G., and Tausch, S.H., 2022. Whole genome sequence analysis of a prototype strain of the novel putative rotavirus species L. *Viruses*, 14(3): 462.

Johne, R., Tausch, S.H., Grützke, J., Falkenhagen, A., Patzina-Mehling, C., Beer, M., Höper, D., and Ulrich, R.G., 2019. Distantly related rotaviruses in common shrews, Germany. *Emerging Infectious Diseases*, 25(12): 2310–2314.

Johne, R., Tausch, S.H., Ulrich, R.G., and Schilling-Loeffler, K., 2023. Genome analysis of the novel putative rotavirus species K. *Virus Research*, 334: 199171.

Jungbauer, A., and Hahn, R., 2009. Ion-exchange chromatography. *Methods in Enzymology,* 463: 349–371.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S.A.A., Ballard, A.J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A.W., Kavukcuoglu, K., Kohli, P., and Hassabis, D., 2021. Highly accurate protein structure prediction with AlphaFold. *Nature,* 596(7873): 583-589.

Kanoh, S., Shiraki, K., Wada, M., Tanaka, T., Kitamura, M., Kato, K., and Hirano, A., 2023. Chromatographic purification of histidine-tagged proteins using zirconia particles modified with phosphate groups. *Journal of Chromatography A, 1703*: 464112.

Karampatsas, K., Osborne, L., Seah, M., Tong, C.Y.W., and Prendergast, A.J., 2018. Clinical characteristics and complications of rotavirus gastroenteritis in children in east London: A retrospective case-control study. *PLoS One*, 13: e0194009.

Kelly, S.M., Jess, T.J., and Price, N.C., 2005. How to study proteins by circular dichroism. Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics, 1751(2): 119–139.

Krimm, S., and Bandekar, J., 1986. Vibrational Spectroscopy and Conformation of Peptides, Polypeptides, and Proteins. *Advances in Protein Chemistry*, 38: 181–364.

Kim, G., Weiss, S.J., and Levine, R.L., 2013. Methionine oxidation and reduction in proteins. *Biochimica et Biophysica Acta (BBA) - General Subjects, 1840*(2): 901-5.

Kirkwood, C.D., Ma, L., Carey, M.E., and Steele, D., 2019. The rotavirus vaccine development pipeline. *Vaccine,* 37(50): 7328–7335.

Kgokolo, S.M., Parbhoo, N., and Gildenhuys, S., 2017. Optimization of purification and characterisation of over-expressed rotavirus capsid protein VP6. M.Sc. (Life Sciences) dissertation. University of South Africa, Pretoria.

Kong, J., and Yu, S., 2007. Fourier transform infrared spectroscopic analysis of protein secondary structures. *Acta Biochimica et Biophysica Sinica, 39*(8): 549–559.

Kumar, N., Malik, Y.S., Kumar, S., Sharma, K., Sircar, S., Saurabh, S., Gulati, B.R., Singh, N., Singh, A.K., Joshi, V.G., Banyai, K., and Dhama, K., 2016. Peptide-recombinant VP6 protein based enzyme immunoassay for the detection of group A rotaviruses in multiple host species. *PLoS ONE, 11*(7): e0159027.

Kuri, P.R., and Goswami, P., 2024. Unravelling aggregation propensity of rotavirus A VP6 expressed as E. coli inclusion bodies through in silico prediction. *Scientific Reports, 14*(1): 21464.

Kuriata, A., Gierut, A.M., Oleniecki, T., Ciemny, M.P., Kołinski, A., Kurcinski, M., and Kmiecik, S., 2019. Aggrescan3D (A3D) 2.0: Prediction and engineering of protein solubility. *Nucleic Acids Research, 47*(W1): W300–W307.

Lakowicz, J.R., 2006. Principles of fluorescence spectroscopy. New York, NY: Springer, 3.

Lawton, J.A., Estes, M.K., and Prasad, B.V.V., 2000. Mechanism of genome transcription in segmented dsRNA viruses. *Adv. Virus Res,* 55: 185–229.

Li, Z., Cui, K., Huang, K., Liu, F., Shi, D., and Liu, Q., 2019. Self-assembling rotavirus VP6 nanoparticle vaccines expressed in *Escherichia coli* elicit systemic and mucosal responses in mice. *Protein and Peptide Letters, 26*(12): 904–909.

Li, R., Nagai, Y., and Nagai, M., 2000. Changes of tyrosine and tryptophan residues in human hemoglobin by oxygen binding: near- and far-UV circular dichroism of isolated chains and recombined hemoglobin. *J Inorg Biochem,* 82 (1-4): 93-101.

Long, C.P., and McDonald, S.M., 2017. Rotavirus genome replication: *Some assembly required*, *13*: e1006242.

López-Monterrubio, D.I., Lobato-Calleros, C., Alvarez-Ramirez, J., and Vernon-Carter, E.J., 2020. Huauzontle (Chenopodium nuttalliae Saff.) protein: Composition, structure, physicochemical and functional properties. *Food Hydrocoll*, *108*: 106043.

Ludert, J.E., Ruiz, M.C., Hidalgo, C., and Liprandi, F., 2002. Antibodies to Rotavirus Outer Capsid Glycoprotein VP7 Neutralize Infectivity by Inhibiting Virion Decapsidation. *Journal of Virology,* 76: 6643–6651.

Madigan, M.T., Beder, K.S., Buckley, D.H., Sattley, W.M, and Stahl, D.A., 2015. *Brock Biology of Microorganism*. Boston: Pearson Edution, 14.

Mahler, H.C., Friess, W., Grauschopf, S.U., and Kiese, S., 2009. Protein aggregation: Pathways, induction factors and analysis. *J Pharm Sci*, 98: 2909-2934.

Mapaseka, S.L., Dewar, J.B., van der Merwe, L., Geyer, A., Tumbo, J., Zweygarth, M., Bos, P., Esona, M.D., Steele, A.D., Sommerfelt, H., 2010. Prospective hospital-based surveillance to estimate rotavirus disease burden in the Gauteng and Northwest Province of South Africa during 2003–2005. *J Infect Dis*, 20: S131–8.

Mathieu, M., Petitpas, I., Navaza, J., Lepault, J., Kohli, E., Pothier, P., Prasad, B.V.V., Cohen, J., and Rey, F.A., 2001. Atomic structure of the major capsid protein of rotavirus: implications for the architecture of the virion. *The EMBO Journal,* 20(7): 1485–1497.

Matthijnssens, J., Otto, P.H., Ciarlet, M., Desselberger, U., Van Ranst, M., and Johne, R., 2012. VP6-sequence-based cutoff values as a criterion for rotavirus species demarcation. *Arch Virol*, 157(6):1177–1182.

Mihalov-Kovács, E., Gellért, Á., Marton, S., Farkas, S.L., Fehér, E., Oldal, M., Jakab, F., Martella, V., and Bányai, K., 2015. Candidate new rotavirus species in sheltered dogs, Hungary. *Emerg Infect Dis,* 21(4): 660–663.

Miles, A.J., and Wallace, B.A., 2016. Circular dichroism spectroscopy of membrane proteins. *Chem Soc Rev,* 45(18): 4859–72.

Miller, L.M., Bourassa, M.W., and Smith, R.J., 2013. FTIR spectroscopic imaging of protein aggregation in living cells. *Biochimica et Biophysica Acta (BBA) – Biomembranes, 1828*(10): 2339–2346.

Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M., 2022. ColabFold: making protein folding accessible to all. Nature Methods, 19: 679–682.

Mirdita, M., Steinegger, M., and Söding, J., 2019. MMseqs2 desktop and local web server app for fast, interactive sequence searches. *Bioinformatics,* 35(16): 2856–2858.

Mital, S., Christie, G., and Dikicioglu, D., 2021. Recombinant expression of insoluble enzymes in *Escherichia coli*: a systematic review of experimental design and its manufacturing implications. *Microbial Cell Factories*, 20(208).

Moore, S., Pickens, A., Rodriguez, J.L., Marsee, J.D., and Miller, J.M., 2020. Fluorescence methods applied to the description of urea-dependent YME1L protease unfolding. *Biomolecules, 10*(4): 656.

Murphy, B.M., D'Antonio, J., Manning, M.C., and Al-Azzam, W., 2014. Use of the amide II infrared band of proteins for secondary structure determination and

comparability of higher order structure. *Current Pharmaceutical Biotechnology. 15*(9): 880–889.

Mwenda, J.M., Parashar, U.D., Cohen, A.L., Tate, J.E., 2018. Impact of rotavirus vaccines in sub-Saharan African countries. *Vaccine,* 36: 7119–23.

Nowakowski, A.B., Wobig, W.J., and Petering, D.H., 2014. Native SDS-PAGE: High resolution electrophoretic separation of proteins with retention of native properties including bound metal ions. *Metallomics*, *6*(5): 1068–1078.

Omatola, C.A., Ogunsakin, R.E., Onoja, A.B., Okolo, M.L.O., Abraham-Oyiguh, J., Mofolorunsho, K.C., Akoh, P.Q., Adejo, O.P., Idakwo, J., Okeme, T.O., Muhammed, D., Adaji, D.M., Samson, S.O., Aminu, R.F., Akor, M.E., Edegbo, E., and Adamu, A.M., 2024. Enteropathogenic viruses associated with acute gastroenteritis among African children under 5 years of age: A systematic review and meta-analysis. *Journal of Infection, 88*(6): 106169.

Pace, C.N., 1986. The stability of globular proteins. *The Journal of Biological Chemistry*, 261(6): 2551-2561.

Pace, C.N., Vajdos, F., Fee, L., Grimsley, G., and Gray, T., 1995. How to measure and predict the molar absorption coefficient of a protein. *Protein Sci*, 11: 2411-2423.

Palmer, I., and Wingfield, P.T., 2012. Preparation and extraction of insoluble (inclusion-body) proteins. *Current Protocols in Protein Science*, 6: 6.3.1–6.3.20

Papa, G., Bokodavka, A., and Desselberger, U., 2021. Viroplasm: Assembly and Function of Rotavirus Replication Factories. *Viruses*,13(7): 1349.

Parbhoo, N., Dewar, J.B., and Gildenhuys, S., 2016. Sequence analysis and structural implications of rotavirus capsid proteins. *Acta Virologica,* 60: 260-270.

Patra, A.K., Mukhopadhyay, R., Mukhija, R., Krishnan, A., Garg, L.C., Panda, A.K., 2000. Optimization of inclusion body solubilization and renaturation of recombinant human growth hormone from *Escherichia coli*. *Protein Expr Purif*,18: 182–92.

Patton, J.T., Jones, M.T., Kalbach, A.N., He, Y.W., and Xiaobo, J., 1997. Rotavirus RNA polymerase requires the core shell protein to synthesize the double-stranded RNA genome. *Journal of Virology*, 71(12): 9618–9626.

Pruitt, K.D., Tatusova, T., Klimke, W., and Maglott, D.R., 2008. NCBI Reference Sequences: current status, policy and new initiatives. *Nucleic Acids Research*, 37: D32–D36.

Pujols, J., Peña-Díaz, S., and Ventura, S., 2018. AGGRESCAN3D: Toward the Prediction of the Aggregation Propensities of Protein Structures. In: Rüdiger S, editor. *Protein Aggregation: Methods and Protocols*. Methods in Molecular Biology, 1762: 427–443.

Qi, X., Sun, Y., and Xiong, S., 2015. A single freeze-thawing cycle for highly efficient solubilisation of inclusion body proteins and its refolding into bioactive form, *Microb Cell Fact.* 14: 24.

Ready, K.F.M., and Sabara, M., 1987. *In vitro* assembly of bovine rotavirus nucleocapsid protein. *Virology,* 157: 189-198.

Robichon, C., Luo, J., Causey, T.B., Benner, J.S., Samuelson, J.C., 2011. Engineering *Escherichia coli* BL21(DE3) Derivative Strains to Minimize *E. Coli* Protein Contamination after Purification by Immobilized Metal Affinity Chromatography, 77(13): 4634-4646.

Rolfe, M.D., Rice, C.J., Lucchini, S., Pin, C., Thompson, A., Cameron, A.D., Alston, M., Stringer, A.M., Betts, R.P., and Hinton, J.C., 2012. Lag phase is a distinct growth phase that prepares bacteria for exponential growth and involves transient metal accumulation. *Journal of Bacteriology*, *194*(3): 686–701.

Rossi, A.M., and Taylor, C.W., 2011. Analysis of protein–ligand interactions by fluorescence polarization. *Nature Protocols, 6*(3): 365–387.

ROTA Council. Available online: https://preventrotavirus.org/vaccine-introduction/global-introduction-status (accessed on 20 January 2022).

Rubick, R., and Kreusch, S., 2018. Desalting, rebuffering, renaturation – Dialysis for optimized sample preparation. *scienova GmbH,* 1-10.

Ruiz, M.C., Leon, T., and Diaz, Y., 2009. Molecular Biology of Rotavirus Entry and Replication. *The Scientific World JOURNAL*. 9: 1476–1497.

Russell, B., and Gildenhuys, S., 2018. Solubilisation and purification of recombinant bluetongue virus VP7 expressed in a bacterial system. *Protein Expr Purif,*147: 85-93.

Sadat, A., and Joye, I.J., 2020. Peak fitting applied to Fourier transform infrared and Raman spectroscopic analysis of proteins. *Sciences*, 10: 5918.

Sandt, C., 2024. Infrared spectroscopy of proteins: A practical approach. *Journal of Molecular Structure*, 1300: 1–12.

Seheri, L.M., Page, N.A., Mawela, M.P.B., Mphahlele, M.J., and Steele, A.D., 2012. Rotavirus vaccination within the South African Expanded Programme on Immunisation. *Vaccine*, 30(3): C14–C20.

Settembre, E.C., Chen, J.Z., Dormitzer, P.R., Grigorieff, N., Harrison, S.C., 2011. Atomic model of an infectious rotavirus particle. *EMBO J*, 30: 408–416.

Sharma, B., Bykov, S.V., and Asher, S.A., 2008. UV resonance Raman investigation of electronic transitions in α-helical and polyproline II-like conformations. *The Journal of Physical Chemistry B, 112*(37): 11762–11769.

Shevchenko, A., Tomas, H., Havlis, J., Olsen, J.V., and Mann, M., 2006. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nature Protocols*, **1**(6): 2856–2860.

Shoja, A., Jalilvand, S., Latif, T., and Roohvand, F., 2022. Rotavirus VP6: involvement in immunogenicity, adjuvant activity, and use as a vector for heterologous peptides, drug delivery, and production of nano biomaterials. *Arch. Virol,* 167: 1013-1023.

Singh, A., Upadhyay, V., Upadhyay, A.K., Mohan Singh, S., and Panda, A.K., 2015. Protein recovery from inclusion bodies of Escherichia coli using mild solubilization process*. Microbial Cell Factories*, 14: 41.

Siligardi, G., Hughes, C.S., and Hussain, R., 2018. Characterisation of sensor kinase by CD spectroscopy: Golden rules and tips. *Biochemical Society Transactions, 46*(6): 1627–1642.

Smok, B., Zieniewicz-Cieślik, K., Smukalska, E., and Pawłowska, M., 2016. Acute diarrhoea induced by rotavirus in children hospitalized in Provincial Hospital for Infectious Diseases in Bydgoszcz in 2014 year. *Przegląd Epidemiologiczny*, 70(3): 462–470.

Sormanni, P., Aprile F.A., Vendruscolo M., 2015. Rational design of antibodies targeting specific epitopes within intrinsically disordered proteins. *Proc Natl Acad Sci USA,*112(32): 9902–9907.

Sosnick, T.R., 2023. *Denaturant Concentration Calculator ([GdmCl], [Urea] calculator)*. The Sosnick Group, University of Chicago.

Srour, B., Bruechert, S., Andrade, S.L.A., and Hellwig, P., 2017. Secondary structure determination by means of ATR-FTIR spectroscopy. In *Methods in Molecular Biology*,1635: 195–203).

Steele, A.D., and Groome, M.J., 2019. Measuring rotavirus vaccine impact in sub-Saharan Africa. *Clinical Infectious Diseases, 70*(11): 2314–2316.

Steinegger, M., and Soding, J., 2017. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol,* 35: 1026–1028.

Taniguchi, K., and Uraswa, S., 1995. Diversity in rotavirus genome. *Seminars in Virology,* 6(2):123-131.

Tao, F., Han, Q., and Yang, P., 2023. Interface-mediated protein aggregation. *Chemical Communications*, 59(95): 14093-14109.

Teale, F.W.J., and Weber, G., 1957. Ultraviolet fluorescence of the aromatic amino acids. *Biochemical Journal, 65*(3): 476–482.

Teng, Y., Zhao, B., Pan, X., Wen, Y., and Chen, Y., 2014. A New Rotavirus VP6-Based Foreign Epitope Presenting Vector and Immunoreactivity of VP4 Epitope Chimeric Proteins. *Viral Immunol,* 27(3): 96–104.

Tintor, D., Ninković, K., Milosevic, J., and Polovic, N.D., 2024. Gaining insight into protein structure via ATR-FTIR spectroscopy. Vibrational Spectroscopy, 134: 103726.

Trevino, S.R., Scholtz, J.M., and Pace, C.N., 2007. Amino acid contribution to protein solubility: Asp, Glu, and Ser contribute more favorably than the other hydrophilic amino acids in RNase Sa. *Journal of Molecular Biology*, 366(2): 449–460.

Uversky, V.N., Gillespie, J.R., and Fink, A.L., 2000. Why are "natively unfolded" proteins unstructured under physiologic conditions? Proteins, 41: 415–427.

Van Damme, P., Giaquinto, C., Todd, P., Wielen, M.V., 2007. On behalf of the REVEAL study group. Distribution of rotavirus genotypes in Europe 2004-2005. The REVEAL study. Journal of Infectious Diseases, 195: S17-S25.

Vivian, J.T., and Callis, P.R., 2001. Mechanisms of tryptophan fluorescence shifts in proteins. *Biophysical Journal, 80*(5): 2093–2109.

Vesikari, T., Karvonen, A., Puustinen, L., Zeng, S.Q., Szakal, E.D., Delem, A., De Vos, B., 2004. Efficacy of RIX4414 live attenuated human rotavirus vaccine in Finnish infants. *Pediatr. Infect. Dis. J*, *23*: 937–943.

Wang, Y., Zhu, L., Zhu, Z., Liu, M., and Zhao, X., 2024. Effects of different pH levels on the structural and functional properties of proteins of *Phaeodactylum tricornutum*. *Molecules, 29*(13): 3139.

Ward, R.L., and McNeal, M.M., 2010. VP6: A Candidate Rotavirus Vaccine. *The Journal of Infectious Diseases,* 202(1): S101–S107.

Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., and Barton, G.J., 2009. Jalview Version 2-a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, 25(9): 1189-1191.

Wei, Y., Thyparambil, A.A., and Latour, R.A., 2014. Protein helical structure determination using CD spectroscopy for solutions with strong background absorbance from 190 to 230 nm. *Biochimica et Biophysica Acta (BBA) – Proteins and Proteomics*, 1844(11): 1953–1961.

Wen, J.H., He, X.H., Feng, Z.S., Li, D.Y., Tang, J.X., and Lie, H.F., 2023. Cellular Protein Aggregates: Formation, Biological Effects, and Ways of Elimination. *Int. J. Mol*. Sci, 24: 8593.

World Health Organization (WHO)., 2021. *Rotavirus vaccines: WHO position paper – July 2021*. Weekly Epidemiological Record, 96(28): 301–319.

Xu, J., and Zhang, Y., 2010. How significant is a protein structure similarity with TM-score = 0.5? *Bioinformatics*, 26(7): 889–895.

Yang, H., Yang, S., Kong, J., Dong, A., and Yu, S., 2015. Obtaining information about protein secondary structures in aqueous solution using Fourier transform IR spectroscopy. *Nature Protocols, 10*(3): 382–396.

Yang, Z., Zeng, X., Zhao, Y., and Chen, R., 2023. AlphaFold2 and its applications in the fields of biology and medicine. *Signal Transduction and Targeted Therapy,* 8(1): 115.

Yao, X., Li, Z., Su, K., Yao, Z., Ren, X., Liu, H., Zhang, Q., Shen, Q., Wang, Q., Liu, Z., Zhang, W., He, Q., 2012. Expression of rotavirus VP6 protein in silkworm larvae using a baculovirus expression system. *J Virol Methods*, 179(2): 421–425.

Yin, Z., Li, M., Niu, C., Yu, M., Xie, X., Haimiti, G., Guo, W., Shi, J., He, Y., Ding, J., Zhang, F., 2023. Design of multi-epitope vaccine candidate against Brucella type IV secretion system (T4SS). PLoS One,18(8): e0286358.

Yong, P.L., Kim, H., Hwang, H.K., Lee, J.S., Kim, H.Y., Hur, B.K., Ryu, Y.W., An, C.N., and Kim, J.S., 2004. Large-Scale Production of Rotavirus VLP as Vaccine Candidate Using Baculovirus Expression Vector System (BEVS). *J. Microbiol. Biotechnol*, 14(1): 35–40.

Zade, J.K., Kulkarni, P.S., Desai, S.A., Sabale, R.N., Naik, S.P., and Dhere, R.M., 2014. Bovine rotavirus pentavalent vaccine development in India. *Vaccine,* 32: A124–A128.

Zhao, Q., Chen, W., Chen, Y., Zhang, L., Zhang, J., and Zhang, Z., 2011. Self-assembled virus-like particles from rotavirus structural protein VP6 for targeted drug delivery. *Bioconjug Chem,* 22(3): 346–352.

Zhang, Y., and Skolnick, J., 2004. Scoring function for automated assessment of protein structure template quality. Proteins: Structure, Function, and Bioinformatics, 57(4), 702–710.

Zou, Q., Habermann-Rottinghaus, S.M., and Murphy, K.P., 1998. Urea effects on protein stability: Hydrogen bonding and the hydrophobic effect. *Proteins: Structure, Function, and Bioinformatics, 31*(2): 107–115.

Zymo Research., 2021. ZymoPURE Plasmid Miniprep Kit protocol. Zymo Research.