

DECLARATION

I hereby declare that “**Molecular Epidemiology of Group A Rotaviruses (RVA) in Children of the State of Imphal, Manipur, India**” has been submitted to Tezpur University in the Department of Molecular Biology and Biotechnology under the School of Sciences for partial fulfilment for the award of the degree of Doctor of Philosophy in Molecular Biology and Biotechnology. This is an original work carried out by me. Further, I declare that no part of this work has been reproduced elsewhere for award of any degree.

Date: 2/8/2023

Place: Tezpur University, Tezpur



(Yengkhom Damayanti Devi)

Registration no. TZ189793 of 2018

DEPARTMENT OF MOLECULAR BIOLOGY AND BIOTECHNOLOGY

(DBT, DST-FIST, UGC-SAP DRS-II SUPPORTED)



तेजपुरविश्वविद्यालय/ TEZPUR UNIVERSITY

(संसदकेअधिनियमद्वारास्थापितकेन्द्रीयविश्वविद्यालय)

(A Central University established by an Act of Parliament)

तेजपुर-784028 :: असम/ TEZPUR-784028 :: ASSAM

(सर्वोत्तमविश्वविद्यालयकेलिएकुलाध्यक्षपुरस्कार, 2016 औरभारतके 100श्रेष्ठउच्चशिक्षणसंस्थानोंमेंपंचमस्थानप्राप्तविश्वविद्यालय)

(Awardee of Visitor's Best University Award, 2016 and 5th among India's Top 100 Universities, MHRD-NIRF Ranking, 2016)

CERTIFICATE OF THE SUPERVISOR

This is to certify that the thesis entitled “**Molecular epidemiology of group A rotaviruses in children of the State of Imphal, Manipur, India**” submitted to the School of Sciences, Tezpur University in partial fulfillment for the award of the degree of Doctor of Philosophy in Molecular Biology and Biotechnology is a record of research work carried out by **Ms. Yengkhom Damayanti Devi** under my supervision and guidance.

All help received by her from various sources have been duly acknowledged. No part of this thesis has been submitted elsewhere for award of any other degree.

Date: 09/08/2022


(Dr. Nima D Namsa)

Assistant Professor

School: School of Sciences

Place: TEZPUR

Department: Molecular Biology and Biotechnology

Acknowledgements

Throughout my doctoral research work, I received a great deal of support and assistance.

First, I would like to express my sincere gratitude to my supervisor, Dr. Nima D Namsa, whose expertise was invaluable in formulating the research questions and methodology. I am always grateful for the support, guidance, motivation, advice, and encouragement he has given me throughout my Ph.D. journey.

My gratitude to Dr. Chongtham Shyamsunder Singh, Dept of Pediatrics, RIMS Imphal, who is our research collaborator, for his kind support, motivation, encouragement, cooperation, care, and concern always.

I want to extend my gratitude to my DC members; Prof Sashi Baruah, Prof. Robin Doley, and Dr. Sougata Saha for their support, guidance, and advice continuously.

Present Head of Dept, Prof. Robin Doley, past Head of Dept; Prof. S. K. Ray and Prof. A. Ramteke have given immense support and cooperation for which I will remain grateful always.

Further, I thank Vice-Chancellor Sir of Tezpur University for providing the infrastructure to conduct my Ph.D. work.

I also express my heartfelt thanks to my Dept faculties for their valuable suggestions and for helping improve my Ph.D. thesis. Much obliged to Dr. Aditya Kumar, Prof. M. Mandal and Dr. Ramsing Thiyam for their kind help with my Ph.D. work.

I am indebted to the department's non-teaching staff, Mrs. Pranita S Talukdar, Dr. K. K. Hazarika, Dr. N. Bordoloi, Mr. P. Mudoi, and Mr. Guna Das for their cooperation and support from time to time.

I will remain grateful to Prof. C Durga Rao, IISc for three months training for my research work and MCB lab members, IISc, for supporting me constantly during my stay in Bangalore.

I want to acknowledge my past lab members for their wonderful cooperation, support, and the great time we had at the beginning of my Ph.D. journey; Hemanga Gogoi, Mimi

Adhikary, Bondita Dehingia, Drema Chozom, Chandrima Doley, Nalini Daimary, Sushmita Konwar and also extend my thanks to my present lab members Kasturi and Deepsikha.

I would also like to thank my friends especially MM lab (Pritam Bardhan, Muzamil Rathor, Kuldip da, and Anupama) and RD lab (particularly Arpita Devi and Rafika Yasmin) for their support and care always. Also, my batch-mates Shreaya Das, Munmi Borah, friends particularly Manoj Sharma, Mayuri Borah, Archana Sinha, Bhaswati, Minhaz, Sayani and Sehnaj, my seniors particularly Ruksana Aziz, Simran Kaur, Archana Singh, Anuj Kumar Borah and all the juniors (particularly Chourajit, Upalabdha, Pratyush, Nyumpi, Sushmita, Panchi) of the department have always been always supportive, cooperative, and caring.

In addition, I also express my gratefulness to UGC for funding my fellowship and DBT for the DBT-Twinning project which was part of my work.

Last but not least; I would like to thank my family for their wise counsel, a sympathetic ear, and standing by me always.

Thank you, God, for all your blessings to me. For the strength you give me each day and for all the people around me who make life more meaningful.

| LIST OF FIGURES | | |
|------------------------|--|-----------------|
| Figure No. | Figure title | Page No. |
| Figure 2.1 | Rotavirus genome and structure organization: (A) Migration pattern on on polyacrylamide gel stained with silver nitrate showing the separation of 11 dsRNA segments of group A rotavirus (RVA) genome. The gene segments are numbered on the left and the corresponding encoded proteins are indicated on the right, (B) Electron microscopy image of the rotavirus triple-layered particle (TLP) (Crawford et al. 2017) and (C) Diagrammatic representation of RV structure and its organization. | 10 |
| Figure 2.2 | Life cycle of rotavirus: RVs attach to glycan receptors and co-receptors on the host cell surface, through interaction VP8* domain of VP4. Initial attachment of some animal strains of RV involves sialic acid receptors (SA sensitive RVs) while human RV strains like Wa and DS-1, often attach to and enter cells in a sialic acid-independent manner (SA insensitive RVs) (Yang Liu 2012) where they interact with potential host receptors like histo blood group antigens (HBGA) (Hu et al., 2012; Huang et al., 2012; Ramani et al., 2013). Cell entry starts with RVs interaction with co-receptors, cellular surface molecules, such as heat shock protein (Hsc70), gangliosides (Martínez, M. A., 2013) and integrins (Graham et al., 2003; Zárate et al., 2004; Gutiérrez et al., 2010, Hussain 2015). Depending on the strain, RVs are internalized into cells by clathrin-dependent or clathrin-independent and caveolin-independent endocytic pathways (Arias CF 2015, Diaz-Salinas MA 2013). Gradient in calcium levels triggers the removal of the outer capsid layer (VP7 and VP4), which releases the transcriptionally active double-layered particle (DLP) into the cytoplasm. Viral (+)RNA is used for translation of the viral proteins and also as a template for RNA synthesis during genome replication. The new virions are assembled and RNA is then packaged into new DLPs within the electron dense structures called viroplasm which are mainly composed of NSP2 and NSP5. Triple-layered particles (TLPs) assembly takes place by the coordinated action of NSP4 and VP6. DLPs buds into the endoplasmic reticulum (ER) forming transiently enveloped particle and the outer capsid proteins VP4 and VP7 are added onto the DLPs. The envelope is lost when RV particles acquire the outer layer consisting of VP4 and VP7 (Estes and Greenberg, 2013). Virions are then released from the cells through cell lysis or Golgi-independent non-classical | 16-17 |

| | | |
|------------|---|-------|
| | vesicular transport mechanism in polarized epithelial cells. | |
| Figure 3.1 | Schematic flowchart showing G-P typing of rotavirus positive stool specimens. | 55 |
| Figure 3.2 | Schematic flowchart showing enrolment criteria for collection of stool specimens from RIMS, Imphal, Manipur. | 57 |
| Figure 3.3 | Bar graph showing seasonal distribution of rotaviral gastroenteritis from December 2015-March 2019. | 63 |
| Figure 3.4 | Rotavirus infection trends among hospitalized children <5 years from Imphal, Manipur, India visited RIMS, Imphal during December 2015 to March 2019. | 64 |
| Figure 3.5 | Amplification of the VP7 and VP4 genes by RT-PCR followed by G and P genotyping. A.i. Amplification products corresponding to VP7 (881bp), A.ii. Amplification products corresponding to VP4 (867bp), B.i. G genotyping by semi-nested multiplex PCR, and B.ii. P genotyping by semi-nested multiplex PCR. lane 1 in all the gel images are DNA ladder (Thermo Scientific-SM0333, SM1163 & SM1333). | 67-68 |
| Figure 3.6 | Frequency of rotavirus G and P genotype distribution isolated from stool specimens positive for rotavirus from children <5 years hospitalized for acute watery diarrhoea (AWD) during December 2015 – March 2019. A.i. Distribution of rotavirus G types strains in the region, A.ii. Distribution of rotavirus mix G type strains, B.i. Distribution of rotavirus P types in the region, and B.ii. Distribution of rotavirus mix P type strains. | 70 |
| Figure 3.7 | Frequency of rotavirus G and P genotype distribution isolated from stool specimens positive for rotavirus from children <5 years hospitalized for acute watery diarrhoea (AWD) during December 2015 – March 2019. A.i. Distribution of rotavirus G types strains in the region, A.ii. Distribution of rotavirus mix G type strains, B.i. Distribution of rotavirus P types in the region, and B.ii. Distribution of rotavirus mix P type strains. | 71 |
| Figure 3.8 | Year-wise distribution of rotavirus genotypes in < 5 years children admitted with acute gastroenteritis, December 2015 – March 2019. | 73 |
| Figure 3.9 | Phylogenetic tree based on the VP7 nucleotide sequence of human rotavirus A strains. The evolutionary history was inferred by using the Maximum Likelihood method and General Time Reversible model. The bootstrap consensus tree inferred from 1000 replicates were taken to represent the evolutionary history of the taxa analyzed. The percentage of trees in which the associated taxa clustered together is shown next to the branches. All positions containing gaps and missing data were eliminated. Evolutionary analyses were conducted in MEGA X. The strains | 74-76 |

| | | |
|-------------|--|-------|
| | in this study are shown in red triangles, the vaccine strains in green triangles and genotypes on the right side. A. G1 strains, B. G2 strains, C. G3 strains and D. G9 strains. | |
| Figure 3.10 | Phylogenetic tree based on the VP4 and VP6 nucleotide sequence of human rotavirus A strains. The evolutionary history was inferred by using the Maximum Likelihood method and General Time Reversible model. The bootstrap consensus tree inferred from 1000 replicates were taken to represent the evolutionary history of the taxa analyzed. The percentage of trees in which the associated taxa clustered together is shown next to the branches. All positions containing gaps and missing data were eliminated. Evolutionary analyses were conducted in MEGA X. The strains in this study are shown in red triangles, the vaccine strains in green triangles and genotypes on the right side. A. P[6] strains, B. P[8] strains and C. VP6. | 77-78 |
| Figure 3.11 | A representative photograph showing the detection of rotavirus, adenovirus, astrovirus and norovirus using CerTest Combo card test (CerTestBiotec). Top green line represents control line and bottom red line indicates presence of the virus under test. (b) Bar graph showing no. of hospitalized individuals infected with rotavirus and co-infection of rotavirus with other three viruses Adenovirus, Astrovirus and Norovirus. | 80 |
| Figure 3.12 | Seasonal prevalence of rotavirus and other common enteric virus (Adenovirus, astrovirus and norovirus) gastroenteritis from December 2015-March 2019. | 81 |
| Figure 3.13 | Detection of rotavirus, adenovirus, astrovirus and norovirus. A. Detection of other enteric viruses by Rota+Adeno+Astro+Noro EIA Combo Card (Cat No. RG862001V) where the green line shows control line and red line shows test line; Agarose gel image for B. adenovirus fibre partial gene (540 bp), C. norovirus genogroup II capsid gene, and D. astrovirus capsid gene PCR product (449 bp). lane 1 in all the gel images are DNA ladder (Thermo Scientific-SM0333 & Axygen® M-DNA-100BP). | 82-83 |
| Figure 3.14 | Phylogenetic tree based on the Adenovirus fiber gene sequence. The evolutionary history was inferred by using the Maximum Likelihood method and General Time Reversible model. The bootstrap consensus tree inferred from 1000 replicates were taken to represent the evolutionary history of the taxa analyzed. The percentage of trees in which the associated taxa clustered together is shown next to the branches. All positions containing gaps and missing data were eliminated. Evolutionary analyses were conducted in MEGA X. | 84 |
| Figure 4.1 | Mutation effect analysis of three amino acid substitutions by | 106 |

| | | |
|------------|---|-----|
| | PROVEAN tool. A. Mutation sites, B. Superimposition image of VP7 protein modelled structure of Rotavac strain and circulating G9 strain. Inset table is mutation effect analysis result; position 87 where it changes from Isoleucine (I) to Threonine (T), position 100 Glycine (G) substituted by Aspartate (D) and position 145 where it changes from Asparagine (N) to Asparatate (D). | |
| Figure 4.2 | Schematic diagram of interactions between VP7 protein of Rotavac strain and Fab region of IgG and interacting residues. Interacting chains (A; VP7, H; Heavy chain, L; Light chain) are joined by coloured lines, each representing a different type of interaction, as per the key above. The area of each circle is proportional to the surface area of the corresponding protein chain. The extent of the interface region on each chain is represented by a coloured wedge whose colour corresponds to the colour of the other chain and whose size signifies the interface surface area. Statistics for all the interfaces are given as inset table. B. Residue interactions across interface Coloured by residue type. The number of H-bond lines between any two residues indicates the number of potential hydrogen bonds between them. For non-bonded contacts, which can be plentiful, the width of the striped line is proportional to the number of atomic contacts. | 107 |
| Figure 4.3 | Schematic diagram of interactions between VP7 protein of circulating G9 strain and Fab region of IgG and interacting residues. Interacting chains (A; VP7, H; Heavy chain, L; Light chain) are joined by coloured lines, each representing a different type of interaction, as per the key above. The area of each circle is proportional to the surface area of the corresponding protein chain. The extent of the interface region on each chain is represented by a coloured wedge whose colour corresponds to the colour of the other chain and whose size signifies the interface surface area. Statistics for all the interfaces are given as inset table. B. Residue interactions across interface Coloured by residue type. The number of H-bond lines between any two residues indicates the number of potential hydrogen bonds between them. For non-bonded contacts, which can be plentiful, the width of the striped line is proportional to the number of atomic contacts. | 108 |
| Figure 5.1 | Laboratory strategy to identify unusual and untypable rotavirus strains. | 121 |
| Figure 5.2 | Workflow for preparation of samples for WGS by Illumina. | 122 |
| Figure 5.3 | Workflow for analysis of raw sequence reads from Illumina | 125 |

| | | |
|------------|---|---------|
| | Novaseq 6000. | |
| Figure 5.4 | Characterization of dsRNA by RNA-PAGE. A. Electrophoretic of 11 segments of RV dsRNA for Group A rotaviruses, B. Electrophoretic of 11 segments of RV dsRNA for Group B rotaviruses. | 128 |
| Figure 5.5 | Detection of Group B and C rotaviruses in VP6-ELISA negative specimens by PCR. Inset table; Primer for Group B rotavirus genes (VP7, VP4 and NSP2) and Group C rotavirus genes (VP7, VP6 and VP4). | 129 |
| Figure 5.6 | A. Patient clinical details of rotavirus isolate RM251122016 qualified for WGS. B. Analysis of RVA RM251122016 dsRNA by RNA-PAGE. | 130 |
| Figure 5.7 | Circos plot of Rotavirus Whole Genome data of RM251122016. Outermost circle represents contigs with their corresponding length in base pair and identified genes. From outer to innermost circles depict GC percentage, query coverage, identity, no. of mapped reads and coverage(X) respectively. Values for each of the features are plotted as ranges from red (high values) to blue (lower values). | 131 |
| Figure 5.8 | Phylogenetic analysis of 11 gene segments of rotavirus isolate RM251122016. A. VP7, B. VP4, C. VP6, D. VP1, E. VP2, F. VP3, G. NSP1, H. NSP2, I. NSP3, J. NSP4 and K. NSP5. | 136-142 |
| Figure 6.1 | Schematic representation of an in-silico strategy followed for prediction of B- and T-cell epitope using proteome of rotavirus. | 158 |
| Figure 6.2 | Localization of selected linear B-cell epitope in native rotavirus protein. A. VP2; 189-AVENKNSRDAGK-200, B. VP3; 238-TIKLKQERWLGGK-249, C. VP4; 208-IPRSEESKCTEYI-220, D. VP4; 241-RDVIHYRAQANED-253, E. VP4; 262-WKEMQYNRDI- 271, F. VP4; 657-PDIVTEASEKF-667, G. VP6; 9-KTLKDARDKIVEG-21, H. VP6; 139-WNLQNRRTQTG-149, I. VP6; 373-NYSPSREDNLQR-384, J. VP7; 308-QVMSKRSRSLNSA- 320, K. NSP2; 267-QNWYAFTSSMKQGNT-281, L. NSP3; 108-LSSKGIDQKMRVL-120, M. NSP4; 117-TTREIEQVELLK-128, N. NSP5; 170-KCKNCKYKKKYFAL-183, O. VP6 (Group A); 74-ISTDDYDDMRSGI-86, P. VP6 (Group B); 197-GMDSEHRFTVELKTR-211, Q. VP6 (Group C); 93-EAVCDDEIVREA-104, R. VP6 (Group C); 143-SRRENVPVYKYKNPM-156. | 168 |
| Figure 6.3 | Structure modeling of rotavirus proteins using I-TASSER. A. VP2, B. VP3, C. VP4, D. VP7, E. NSP1, F. NSP3, G. NSP4 and H. NSP5. | 169 |
| Figure 6.4 | Summary of rotavirus protein-derived B- and T-cell epitopes. | 191 |

| | | |
|------------|--|---------|
| | Heat map showing the distribution of (A) linear (continuous) B-cell epitopes, (B) HLA-class I and (C) HLA-class II epitopes across the structural and non-structural protein sequences of rotavirus. Strong binding affinity epitopes with <0.5% rank and 2% rank, to HLA class I and class II, respectively, for each HLA molecule are represented here. Red color represents likely antigenic epitopes that were predicted using the methods described in Figure 6.1. | |
| Figure 6.5 | Schematic diagram of multi-epitope chimeric constructs. The multi-epitope constructs sequence consisting adjuvant followed by T- and B-cell epitope. Adjuvant and CTL epitope has been joined by EAAAK linker, whereas the AAY, KK and GGGGS linkers were used to join the CTL, HTL and linear/conformational B-cell epitopes, respectively. A. Construct 1 (VP6A/B/C), B. Construct 2 (VP6/4/7), C. Construct 3 (VP2/3/4/6/7), D. Construct 4 (NSP2/3/4/5) and E. Construct 5 (VP2/3/4/6/7-NSP2/3/4/5), F. Construct 6 (VP6A/B/C-B), G. Construct 7 (VP4/6/7-B), H. Construct 8 (VP4/A), I. Construct 9 (VP6/A) and J. Construct 10 (VP7/A); (BL-Linear B-cell epitope, BC- Conformational B-cell epitope). A/B/C: VP6 sequence of group A, B and C rotaviruses; A: group A rotavirus; B: B-cell epitopes (Both linear and conformational). | 201 |
| Figure 6.6 | Graphical representation of secondary structure obtained for the multi-epitope constructs using PSIPRED server. A. Construct 1, 52.2% helix, 3.0% sheet and 44.8% coil, B. Construct 2, 10.43% helix, 9.71% sheet and 79.86% coil, C. Construct 6, 28.5% helix, 4.0% sheet and 67.5% coil and D. Construct 7, 19.2% helix, 4.8% sheet and 76.0% coil, E. Construct 3, 36.2% helix, 11.0% sheet and 52.8% coil, F. Construct 4, 25.39% helix, 11.92% sheet and 62.69% coil, G. Construct 5, 34.0% Helix, 3.7% Sheet and 62.3% Coil, H. Construct 8, 11.5% helix, 32.7% sheet and 55.8% coil, I. Construct 9, 40.2 % helix, 2% sheet and 57.8 % coil and J. Construct 10, 59.4 % helix, 11.6 % sheet and 29 % coil. | 205-207 |
| Figure 6.7 | Tertiary structure modeling of multi-epitope constructs. Cyan color represents CTL epitopes, orange represents HTL epitopes, blue represents linear B- cell epitopes and conformational B-cell epitope is highlighted with green. A. Construct 1; B. Construct 2; C. Construct 3; D. Construct 4; E. Construct 5; F. Construct 6; G. Construct 7; H. Construct 8; I. Construct 9; and J. Construct 10. | 208 |
| Figure 6.8 | Molecular dynamics simulation study of final multi-epitope constructs representing root mean square deviation. A. | 209 |

| | | |
|-------------|--|-----|
| | Simulation was carried out for time duration of 20 ns for construct 1, 2, 6 and 7, B. Simulation for construct 5 was performed for the time duration of 40 ns, C. Simulation for 20ns for constructs 3,4,8,9 and 10. | |
| Figure 6.9 | Structure prediction and validation of final multi-epitope constructs. Ramachandran plot analysis of the simulated structures. Summary of residues in favored, allowed and in outlier part is provided in Table 4. | 210 |
| Figure 6.10 | Surface accessibility of linkers in the final multi-epitope constructs. A. Construct 1; B. Construct 2; C. Construct 3; D. Construct 4; E. Construct 5; F. Construct 6; G. Construct 7; H. Construct 8; I. Construct 9; and J. Construct 10. Blue color represents AAY linker, cyan represents KK linker and GGGGS is represented by red color. | 212 |
| Figure 6.11 | Conformational B-cell epitopes prediction for the final multi-epitope constructs by Ellipro. A. Construct 1, B. Construct 2, C. Construct 3, D. Construct 4, E. Construct 5, F. Construct 6, G. Construct 7, H. Construct 8, I. Construct 9 and J. Construct 10. The epitopes are represented as colored spheres in the final vaccine model where each color represents one epitope. | 213 |
| Figure 6.12 | Docked complex of multi-subunit vaccine constructs with integrin receptor. A. Construct1 interaction with α IIb β 3 B. Construct 1 with α V β 3 C. Construct 2 with α IIb β 3 D. Construct 2 with α V β 3 E. Construct 6 with α IIb β 3 F. Construct 6 with α V β 3 G. Construct 7 with α IIb β 3 H. Construct 7 with α V β 3. Integrin receptor chain A and B has been shown in cyan and silver color, respectively, whereas magenta color represents the multiepitope vaccine constructs in the docked complex. | 221 |
| Figure 6.13 | Cloning, expression, and affinity purification of four chimeric constructs in E. coli. A) Confirmation of recombinant clones using PCR. Construct 1 (expected gene size 693 bp); Construct 2 (expected gene size 837 bp with 201 bp from vector sequence due to use of T7 forward primer), Construct 6 (expected gene size 603 bp) and Construct 7 (expected gene size 690 bp) were synthesized and cloned into champion pET directional TOPO expression vector (pET100/D-TOPO). C: Negative control without template DNA; M: GeneRuler 1 kb DNA ladder (SM0311, Thermo Scientific). B) pET100/D-TOPO vector map, C) SDS-PAGE analysis showing the expression of recombinant chimeric proteins induced with IPTG (200 μ M) at 18°C, 25°C and 37°C induction temperature. Construct 1 (expected size 28.9 kDa including tag); Construct 2 (expected size 34 kDa including tag); Construct 6 (expected size 28.9 kDa including tag); | 226 |

| | | |
|-------------|---|---------|
| | Construct 7 (expected size 34 kDa including tag); UI: uninduced E. coli whole cell lysates. | |
| Figure 6.14 | <p>A. Analysis of recombinant protein in soluble and insoluble fractions of bacterial whole cell lysate. SDS-PAGE analysis showing the expressed recombinant chimeric proteins induced with IPTG (200 μM) at 25°C in soluble fraction. Construct 1 (expected size 28.9 kDa including tag); Construct 2 (expected size 34 kDa including tag); Construct 6 (expected size 28.9 kDa including tag); Construct 7 (expected size 34 kDa including tag); M: Protein Marker, UI: uninduced E. coli whole cell lysates, SO: Soluble fraction, IN/S: Insoluble Supernatant, IN/P: Insoluble Pellet, B. Purification of soluble protein by Ni-NTA affinity chromatography and confirmation by SDS-PAGE followed by silver staining. Construct 1 (expected size 28.9 kDa including tag); Construct 2 (expected size 34 kDa including tag); Construct 6 (expected size 28.9 kDa including tag); Construct 7 (expected size 34 kDa including tag); M: Protein Marker, FT: Flow Through, W1: First Wash, W2: Second Wash, E1, E2 and E3: First, second and third elute. Inset table: Expected size of construct 1 & 2 and concentration of protein estimated (ng/μl) for 100 μl elution, C. Purification of soluble protein by Ni-NTA affinity chromatography and confirmation by SDS-PAGE followed by silver staining. Construct 1 (expected size 28.9 kDa including tag); Construct 2 (expected size 34 kDa including tag); Construct 6 (expected size 28.9 kDa including tag); Construct 7 (expected size 34 kDa including tag); M: Protein Marker, FT: Flow Through, W1: First Wash, W2: Second Wash, E1, E2 and E3: First, second and third elute. Inset table: Expected size of construct 6 & 7 and concentration of protein estimated (ng/μl) for 100 μl elution.</p> | 227-228 |
| Figure 6.15 | Silver stained-SDS-PAGE gel electrophoresis showing the purity of multi-epitope antigens. Different concentrations of BSA were loaded to determine the approximate concentration of purified proteins estimated using Bradford assay. M: Prestained protein ladder (Cat. 26616, Thermo Scientific). | 229 |

| LIST OF TABLES | | |
|-----------------------|---|-----------------|
| Table No. | Table title | Page No. |
| Table 2.1 | Rotavirus structural and non-structural proteins. | 11 |
| Table 2.2a | Vesikari clinical severity scoring system (VSS). VSS was determined to scale the severity of rotavirus induced acute gastroenteritis based on seven clinical parameters recorded at the hospital | 25 |
| Table 2.2b | Vesikari Clinical Severity Scoring System Severity Rating Scale | 25 |
| Table 3.1 | Numbers and proportions of children enrolled, tested for rotavirus and having rotavirus-positive stool specimens, December 2015 – March 2019 | 61 |
| Table 3.2. | Characteristics of rotavirus-infected and -uninfected children (≤ 5 years) hospitalized with acute gastroenteritis. | 62 |
| Table 3.3 | Consensus and G/P typing primers used for the amplification VP7, VP4, VP6, NSP4 and NSP5 genes. | 65 |
| Table 3.4 | Age wise prevalence of rotavirus in children (≤ 5 years) hospitalized with acute gastroenteritis. | 69 |
| Table 3.5 | Genotype distribution at the surveillance site during the 3 years study period in children ≤ 5 years. | 72 |
| Table 3.6 | Oligonucleotide primers used for PCR amplification of other enteric viruses detected in clinical stool samples; norovirus, adenovirus and astrovirus genes. | 82 |
| Table 3.7 | Enrolment criteria for collection of stool specimens from RIMS, Imphal, Manipur. | 56 |
| Table 4.1 | Summary of VP7 and VP4 lineage, date of isolation, and available patient information for rotavirus strains from Imphal analyzed in this study and rotavirus vaccine strains Rotavac, Rotarix and RotaTeq. | 98 |
| Table 4.2 | The nucleotides and amino acids differences of VP7 and VP4 proteins of Manipur strains with rotavirus vaccine strains. Intragenotype similarities with Rotarix are colored in blue. Intragenotype similarities with RotaTeq are colored in orange. Intragenotype similarities with Rotavac are colored in green. | 99 |
| Table 4.3 | Alignment of antigenic residues in VP7 between the strains contained in Rotarix, RotaTeq and Rotavac and strains circulating in Imphal. Antigenic residues are divided in three epitopes (7-1a, 7-1b, and 7-2). Amino acids that differ between Rotarix, RotaTeq and Rotavac are indicated in boldface. Red colored residues are residues that are different from Rotarix, and Green colored residues are different from the most similar genotype in RotaTeq. Residues colored in Blue are different | 102 |

| | | |
|------------|---|---------|
| | from both Rotarix and RotaTeq, Residues colored in Orange are different from Rotavac. Amino acid changes that have been shown to escape neutralization with monoclonal antibodies (McDonald SM 2009) are indicated with asteric. | |
| Table 4.4 | Alignment of antigenic residues in VP4 (VP8*) between the strains contained in Rotarix, RotaTeq and Rotavac and strains circulating in Imphal. Antigenic residues are divided in four epitopes (8-1, 8-2, 8-3 and 8-4). Amino acids that differ between Rotarix and RotaTeq are indicated in boldface. Orange colored residues are residues that are different from Rotarix, and Green colored residues are different from the most similar genotype in RotaTeq. Residues colored in Blue are different from both Rotarix and RotaTeq. Amino acid changes that have been shown to escape neutralization with monoclonal antibodies (McDonald SM 2009) are indicated with asteric. | 103 |
| Table 5.1 | Assembly and genotyping details of the RVA RM251122016 genome. | 123 |
| Table 5.2 | Surveillance of mutations in antigenic region of VP7 between RM251122016, strains from VP7 tree and vaccine strain. | 126 |
| Table 5.3 | Surveillance of mutations in antigenic region of VP4 between RM251122016, strains from VP4 tree and vaccine strain. | 144 |
| Table 5.4 | Surveillance of mutations in antigenic region of VP6 between RM251122016, strains from VP7 tree and vaccine strain. | 145 |
| Table 5.5 | Surveillance of mutations in antigenic region of VP6 between RM251122016, RVA sequence in NCBI database and vaccine strain | 146 |
| Table 6.1 | Physico-chemical analysis and prediction of antigenicity of rotavirus proteins | 164 |
| Table 6.2 | Summary of linear B-cell epitopes predicted by Bcepred. Four physicochemical properties namely, hydrophilicity (Property 1), flexibility (Property 2), polarity (Property 3) and exposed surface (Property 4) were considered for prediction of B-cell epitopes. The selected epitopes used for designing multi-epitope vaccine constructs are highlighted in bold. | 166-167 |
| Table 6.3 | Summary of predicted conformational B-cell epitopes. Four servers were used based on amino acid sequence dependent, CBTOPE (1) and three structure-dependent tools, Ellipro (2), Discotope (3) and EPSVR (4). The selected epitopes used for designing multi-epitope vaccine constructs are highlighted in bold | 170-171 |
| Table 6.4a | List of HLA class supertype alleles used in this study | 173 |
| Table 6.4b | List of HLA class II supertype alleles used in this study | 174 |
| Table 6.5a | Summary of predicted MHC I binding epitopes. Three different | 175-176 |

| | | |
|------------|---|---------|
| | tools were used and servers that predicted the corresponding epitope are numbered as IEDB Proteasomal cleavage/TAP transport/MHC class I combined predictor (1), nHLAPred (2) and RankPep (3). The selected epitopes used for designing multi-epitope vaccine constructs are highlighted in bold | |
| Table 6.5b | Summary of predicted MHC II binding epitopes. Three different tools were used and servers that predicted the corresponding epitope are numbered as NetMHCpan 3.1 (1), ProPred (2) and RankPep (3). The selected epitopes used for designing multi-epitope vaccine constructs are highlighted in bold | 177-178 |
| Table 6.6a | Molecular docking of predicted CTL epitopes with MHC I complexes. The selected epitopes used for designing multi-epitope vaccine constructs are highlighted in bold | 180-186 |
| Table 6.6b | Molecular docking of predicted HTL epitopes with MHC II complexes. The selected epitopes used for designing multi-epitope vaccine constructs are highlighted in bold | 187-190 |
| Table 6.7 | Rotavirus proteins, total number of epitopes predicted and the immunogenicity/antigenicity/allergenicity as obtained from immune epitope database | 192-195 |
| Table 6.8 | Predicted B- and T-cell epitopes obtained from the immune epitope database. The amino acid sequence of selected epitopes used for design of final multi-subunit chimeric antigen constructs | 198-200 |
| Table 6.9 | Physico-chemical parameter of final multi-epitope constructs. Number of residues, theoretical pI, molecular weight, aliphatic index, and grand average of hydrophobicity (GRAVY) by ProtParam | 203 |
| Table 6.10 | Summary of amino acid residues of vaccine constructs in the energetically favored, allowed and residues in the outlier region as analyzed by physico-chemical parameter of final multi-epitope constructs | 211 |
| Table 6.11 | Verification of conformational B-cell epitopes. The epitopes in the final vaccine constructs were superimposed with their native structure by Phymol and Yale alignment server. The corresponding B-cell epitopes or part of the epitope predicted using four servers are numbered as CBTOPE (1), Ellipro (2), Discotope (3) and EPSVR (4). | 214-216 |
| Table 6.12 | Conformational B-cell epitopes prediction for the final multi-epitope vaccine construct by Ellipro | 217-219 |
| Table 6.13 | A list of interacting residues of docked multi-subunit vaccine constructs with integrin receptor complex | 222 |

ABBREVIATIONS

| | |
|--------|--|
| RV | Rotavirus |
| RVA | Group A rotavirus |
| GE | Gastroenteritis |
| ICTV | International Committee on Taxonomy of Viruses |
| DLP | Double-layer particle |
| TLP | Triple-layer particle |
| dsRNA | Double-stranded RNA |
| VP | Viral proteins |
| NSP | Non-structural proteins |
| WHO | World Health Organization |
| ROTA | Rotavirus Organization of Technical Allies |
| NRSN | National Rotavirus Surveillance Network |
| UIP | Universal Immunization Program |
| NER | North-eastern states |
| EM | Electron Microscopy |
| RdRp | RNA dependent RNA polymerase |
| LRI | Long Range Interactions |
| ADRV-N | Novel Adult Diarrhoea Rotavirus |
| AGMK | African green monkey kidney |
| HBGA | Histo Blood Group Antigens |
| EE | Early Endosomes |
| ME | Maturing Endosomes |
| ILVs | Intraluminal Vesicles |
| LEs | Late Endosomes |
| TCs | Transcription Complexes |
| LDs | Lipid Droplets |
| PABP | Poly (A)-Binding Protein |

| | |
|-------------|---|
| CSF | Cerebrospinal Fluid |
| AGE | Acute Gastroenteritis |
| Native-PAGE | Native Polyacrylamide Gel Electrophoresis |
| E-type | Electropherotypes |
| NGS | Next Generation Sequencing |
| BQR | Brequinar |
| LFM | Leflunomide |
| DHODH | Dihydroorotate Dehydrogenase |
| EE | Environmental Enteropathy |
| OPV | Oral Polio Vaccine |
| DRC | Democratic Republic of the Congo |
| EtBr | Ethidium Bromide |
| MEGA | Molecular Evolutionary Genetics Analysis |
| NCBI | National Center for Biotechnology Information |
| RIMS | Regional Institute of Medical Sciences |
| EIA | Enzyme Immunoassay |
| ELISA | Enzyme Linked Immuno-sorbent Assay |
| RT-PCR | Reverse Transcription Polymerase Chain Reaction |
| PROVEAN | Protein Variation Effect Analyzer |
| I-TASSER | Iterative Threading ASSEMBly Refinement |
| RMSD | Root Mean Square Deviation |
| ViPR | Virus Pathogen Database and Analysis Resource |
| WGS | Whole genome sequencing |
| SVR | Support Vector Regression |
| LINCS | Linear Constant Solver |
| CTL | Cytotoxic T lymphocytes |
| HTL | Helper T-cell |
| RGD | Integrin binding motif |
| MD | Molecular dynamics |
| SA11 | Simian Agent 11 |
| IEDB | Immune Epitope Database |
| IS | Intussusception |