CHAPTER-6

DOMAIN MAPPING OF CHPV NUCLEOCAPSID PROTEIN
6.1 Introduction

A marked feature of negative strand RNA viruses which differentiates them from other viruses is the enwrapped RNA genome. The genomic RNA is always encapsidated by virally encoded nucleocapsid protein to form ribonucleoprotein (RNP) complex. Interactions among the proteins of the RNP complex are essential for functioning of key processes during virus replication and transcription. This RNP complex is unique to negative strand RNA viruses and the distinctive functions performed by RNP of these viruses require some conserved structural motifs in the nucleocapsid protein like, the (5H+3H) motif has been identified to be conserved in N protein of negative strand RNA viruses [163]. N protein of rhabdoviruses is a multifunctional protein that plays a central role in viral RNA transcription and replication and it does so by virtue of its interactions with other viral proteins. The interaction of monomeric N protein with P maintains it in the encapsidation competent soluble (active) form [20, 81] while its association with M protein facilitates the condensation of the enwrapped genome thus giving the virion its characteristic bullet shape [48].

The interactions among Chandipura virus proteins have been revealed and confirmed in previous chapters. Interestingly, N protein was observed to interact with all four viral proteins considered in the study i.e., N, P, M and G. The binding of CHPV N with M and G proteins was identified for the first time while binding with P and self was known earlier. This observation was the basis of further work aimed to identify the interacting residues of N protein involved in its interactions with other viral proteins by domain mapping studies.

Mondal and coworkers have mapped the domains responsible for NN and NP interaction in CHPV by deletion studies. The N-terminal 47 amino acids and 180-264 residues have been shown to be indispensable for N protein oligomerization [19] while the N-terminal 180 aa and the C-terminal 102 aa of N protein are required for binding to P protein in its monomeric and RNA encapsidated state, respectively [88]. NN and NP interactions have also been studied at the domain levels for the related viruses like VSV [40, 164] and RV [165]; however, these studies were based on random deletion and mutation experiments.

In an attempt to examine the specific regions of CHPV N protein responsible for its interactions with N, P, M and G proteins, the structures of CHPV N and M proteins were generated by homology modeling using SWISS-MODEL workspace, G protein using I-TASSER and P protein using a modified \textit{ab-initio} method. The predicted structures were validated by Structure Analysis and Verification Server (SAVES). Using another computational tool, PiSQRD, the nucleocapsid protein was divided into three overlapping
fragments- an N-terminal fragment N1 (Nt arm), central N2 (Nt lobe) and a C-terminal N3 (Ct lobe). The involvement of N protein fragments in the interactions of N with N, P, M and G proteins were predicted using ZDOCK and RDOCK and these bioinformatics based predictions were followed by experimental analysis using yeast two-hybrid and ELISA assays.
6.2 Materials and methods

6.2.1 Structure elucidation of Chandipura virus proteins by homology modeling

CHPV N and M protein structural models were generated using SWISS-MODEL workspace. The templates used for homology modeling of these proteins were Vesicular Stomatitis virus (VSV), Indiana strain N protein complex (PDB ID: 3HHW) and M protein (PDB ID: 1LG7) as both CHPV and VSV belong to the same genus (Vesiculovirus). N and M proteins of both viruses share 71.3% and 28% sequence similarity, respectively. All the corresponding protein sequence alignments were obtained using CLUSTALW with default parameters and were fed into the SWISS-MODEL server. The protein structure of CHPV G protein was generated using I-TASSER [166, 167] by defining VSV pre-fusion form of Glycoprotein complex (PDB ID: 2J6J) as the template. Inter-atomic clashes and bond length errors within G protein structure were removed by using ModRefiner [168]. The tertiary structure obtained was matched with the predicted secondary structure (JPRED and PSSpred) and the transmembrane domain together with the ectodomain were remodeled individually using MODELLER (9v4) by assigning conformation restrictions resulting in the generation of full length CHPV G protein structure.

The generation of CHPV P protein structure using homology modeling and threading method was not possible due to the unavailability of full-length structural homologues for the protein sequence. The templates that exhibited homology to CHPV P protein were VSV P protein oligomerization domain (PDB ID: 2FQM) and C terminal domain (PDB ID: 2K47). The structural model of P protein was thus generated using *ab-initio* method in combination with manual optimization. P protein structure was generated by dividing the protein into four segments, N-terminal domain (residues 1-103), oligomerization domain (residues 104-168), interconnecting domain (residues 169-215), and C terminal domain (residues 216-289). QUARK [169], I-TASSER [166, 167] and SWISS-MODEL workspace were used to model the individual segments in agreement with their secondary structure predictions. Oligomerization and C terminal domains were modeled using SWISS-MODEL workspace using templates available from VSV P protein, while N-terminal and interconnecting domain models were obtained using I-TASSER and QUARK, respectively. The disulfide connectivity was predicted using DiANNA server [170]. Complete structure was generated by assembling the individually modeled segments; oligomerization domain, interconnecting domain, and C terminal domain manually around N terminal domain with several biochemical properties of
the protein as guiding parameters. The complete model was assembled using Swiss-PDB Viewer by reproducing the torsion angles from the individual modeled segments to the extended polypeptide chain (289 aa). The torsion angles were modified wherever needed based on disulfide connectivity and secondary structure prediction details. Close contacts between atoms and geometrical incompatibility were also considered as excluding parameters during the protein modeling. Between each transformation-rotation cycle the protein structure was refined using ModRefiner [168] by keeping both side-chain and backbone atoms completely flexible during structure refinement simulations. All the protein structures prepared for CHPV (N, P, M and G) were subjected to prepare protein protocol of Accelrys Discovery Studio Client 2.55 for the energy minimization, optimizing short and medium sized loop regions and protonating the protein structures. Protein dielectric constant was set to 10 and a pH of 7.5 was used with an ionic strength of 0.150 M for protonation. The disulfide bridges were left intact during the minimization and optimization process. Ramachandran plot, Verify3D and ERRAT were used to estimate the stereo chemical quality.

6.2.2 Decomposing N protein into N1, N2 and N3 fragments

The CHPV N protein structure was subdivided in three fragments (N1-Nt arm, N2- Nt lobe and N3-Ct lobe) using PiSQRD web resource. The server is based on β-Gaussian network elastic model and uses an algorithm introduced by Potestio and co-workers [171]. It subdivides proteins into regions that behave as rigid units in the course of protein structural fluctuations. The regions predicted were altered based on SEQRES database for minimizing the structure assembling bias.

6.2.3 Docking of CHPV N protein

The docking of CHPV N protein with N, P, M and G proteins was performed using the ZDOCK (Accelrys) [172, 173] rigid-body docking method. ZDOCK uses the ZDOCK algorithm and also allows the clustering of poses according to the ligand position thus providing rigid body docking of two protein structures. Additionally, for all the ZDOCK calculations, a non-deterministic FFT optimization was also employed. A cut off distance of 10 Å was used for all the protein pairs to define the interface region between receptor and ligand proteins. The ZDOCK generated docked protein poses were re-ranked on the basis of a more detailed energy function using ZRANK. RDOCK (Accelrys) [174, 175] refinement was performed on the top 100 poses of the filtered ZDOCK output of each interacting pair.
RDOCK uses CHARMm to energy optimize a set of docked protein poses generated by the ZDOCK. The receptor and all the ligand protein molecules were typed with CHARMm Polar H force field prior to the refinement. A distance dependent dielectric constant $4 \pi r$ (where $r$ is the distance) was used during refinement. Scaling factor ($\beta$), that weighs the contribution of the desolvation energy compared to the electrostatic energy in the total RDock energy, was set to 0.9.

$$E_{\text{RDock}} = E_{\text{sol}} + \beta E_{\text{elec2}}$$

Additional RNA binding and oligomerization constraints were also employed as special cases for the determination of N-P conformational binding in the presence of RNA in the system and on the association of other N protein monomeric molecules to form oligomers, respectively. The design of these constraints excluded those amino acid residues which lay in the interface region between CHPV N protein and RNA or between different molecules of CHPV N protein.

### 6.2.4 Cloning of N protein fragments in yeast and bacterial expression vectors

The Nucleocapsid (N) gene boundaries corresponding to N1, N2 and N3 regions were identified and these regions were amplified using specifically designed primers (Table 6.1). The primers were designed to include Nde I and BamH I restriction enzyme sites at their 5’ ends to facilitate cloning in yeast expression vectors pGBK T7 (BD, bait) and pGADT7 (AD, prey) [Clontech]. The primer pairs were designed to amplify N1 (1-90 bp), N2 (61-894 bp) and N3 (693-1260 bp) from nucleocapsid ORF cloned in pET33b vector [20]. The three fragments were amplified by the PCR protocol described in section 3.2.3 followed by their purification using Qiagen PCR purification kit (Section 3.2.4) and then were digested using Nde I and BamH I enzymes (Fermentas). The vectors, pGBK T7 and pGADT7 were also digested with the same enzyme combination and then ligated with digested and purified N fragments to generate the respective BD and AD fusion constructs. The recombinant pGBK T7 vectors, called BD-N1, BD-N2 and BD-N3, encoded the fragments N1 (30 amino acids), N2 (278 amino acids) and N3 (191 amino acids) fused in frame at the C terminus of binding domain of BD vector. The pGADT7 constructs, called AD-N1, AD-N2 and AD-N3, encoded the corresponding fragments fused in frame downstream of activation domain of AD vector.

Fusion constructs for ELISA were also generated by PCR amplification using primers corresponding to the sequences of N1, N2 and N3 regions using N-pET33b as template. The primers were designed to incorporate BamHI and XhoI sites (Table 6.1) for cloning in
pGEX-4T3 vector. The amplified products were purified and digested. The pGEX-4T3 vector was also digested with BamH I and Xho I, purified and ligated with the amplified and digested N gene fragments. The pGEX-4T3 constructs called GST-N1, GST-N2 and GST-N3 contain the three fragments fused in frame with GST tag at the N terminus. BL21 cells carrying GST-N fragment fusions were induced with 1 mM IPTG at 25 °C for four hours. The cells were lysed using IBA lysis buffer and the supernatant fractions were analysed for the presence of N1, N2 and N3 by western blot using anti GST monoclonal antibody. The detailed steps for cloning, protocols for induction, lysis and western blot procedure have been described in Sections 3.2.6, 3.2.7, 3.2.8.1-8.3).

**Table 6.1:** Primers used for cloning N protein N1, N2 and N3 in BD and AD yeast vectors and vector pGEX-4T3

<table>
<thead>
<tr>
<th>S.No.</th>
<th>Construct</th>
<th>Oligo Name</th>
<th>Primer Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>BD-N1 AD-N1</td>
<td>N1 F (BD/AD)</td>
<td>5’GGAAGTGCA CATATGAGTTCTCAAGTATTCTG C3’</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N1 R BamH I (BD/AD)</td>
<td>5’GCTAACAGGAT CCGAGAGAATGCCCTGGGA AAC3’</td>
</tr>
<tr>
<td>2</td>
<td>BD-N2 AD-N2</td>
<td>N2 F (BD/AD)</td>
<td>5’GGAAGTGCA CATATGGAAAGACCCAGT GGA GTTC3’</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N2 R BamH I (BD/AD)</td>
<td>5’GCTAACAGGAT CCGAGATGGAAACTG GGATTTTTGTTG C3’</td>
</tr>
<tr>
<td>3</td>
<td>BD-N3 AD-N3</td>
<td>N3 F (BD/AD)</td>
<td>5’GGAAGTGCA CATATGAC TCTGTCACACCTCCAG3’</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N3 R BamH I (BD/AD)</td>
<td>5’GCTAACAGGAT CCG AGATGGAAACTGGGATTTTTGTTG C3’</td>
</tr>
<tr>
<td>4</td>
<td>GST-N1</td>
<td>N1 F BamH I (pGEX-4T3)</td>
<td>5’GCTAACAGGAT CCG AGATGGAAACTGGGATTTTTGTTG C3’</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N1 R Xho I (pGEX-4T3)</td>
<td>5’GCTAACAGGAT CCG AGATGGAAACTGGGATTTTTGTTG C3’</td>
</tr>
<tr>
<td>5</td>
<td>GST-N2</td>
<td>N2 F BamH I (pGEX-4T3)</td>
<td>5’GCTAACAGGAT CCG AGATGGAAACTGGGATTTTTGTTG C3’</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N2 R Xho I (pGEX-4T3)</td>
<td>5’GCTAACAGGAT CCG AGATGGAAACTGGGATTTTTGTTG C3’</td>
</tr>
<tr>
<td>6</td>
<td>GST-N3</td>
<td>N3 F BamH I (pGEX-4T3)</td>
<td>5’GCTAACAGGAT CCG AGATGGAAACTGGGATTTTTGTTG C3’</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N3 R Xho I (pGEX-4T3)</td>
<td>5’GCTAACAGGAT CCG AGATGGAAACTGGGATTTTTGTTG C3’</td>
</tr>
</tbody>
</table>

Restriction enzyme sites are in bold and the name of enzymes in italics. The vector names are given in parenthesis.

**6.2.5 Yeast transformation, mating and screening protein interactions**

The BD and AD fusion constructs with N fragments (N1, N2 and N3) were transformed individually in *S. cerevisiae* strains Y187 and AH109, respectively, by lithium acetate yeast transformation protocol as explained in Section 4.2.6.3. The BD and AD transformants were plated on SD media lacking amino acids tryptophan and leucine,
respectively. Both BD and AD constructs were also screened for autoactivation of the reporter gene HIS3 on SD media lacking tryptophan and histidine (SD/-Trp/-His) for BD and on SD/-Leu/-His media for AD constructs, respectively. Each of the recombinant BD fusion in Y187 strain was allowed to mate with each prey construct in AH109 yeast strain (protocol detailed in section 4.2.6.8.2). Y187 cells harbouring recombinant BD vectors (BD-N1, BD-N2 and BD-N3) were mated with AH109 cells carrying AD vectors expressing four complete CHPV proteins (N, P, M and G) accounting for a total of 12 mating arrangements. Similarly, the assay was carried out in reverse order i.e., AH109 cells harbouring recombinant AD vectors (AD-N1, AD-N2 and AD-N3) were mated with Y187 cells carrying BD vectors expressing four complete CHPV proteins (N, P, M and G) accounting for another 12 mating pairs. Mated diploids with both bait and prey vectors were selected on SD media lacking tryptophan and leucine (SD/-Trp/-Leu) and tested for protein-protein interaction by plating on SD/-Trp/-Leu/-His media. Concurrently, the mated clones were also screened by α-galactosidase assay by plating on SD/-Trp/-Leu/-His/α-gal media. The development of blue colour in the presence of X α-gal was indicative of positive interaction in this assay. Positive (pGBK7-53+pGADT7-T) and negative (pGBK7-Lam+pGADT7-T) interaction controls were also used in the assay.

6.2.6 Enzyme linked immunosorbent assay (ELISA) for detection of interactions

Streptactin coated microtiter plate (IBA-GmbH) was used to check the interactions among full length CHPV proteins as Strep fusions (Strep-N, Strep-P, Strep-M and Strep-G; soluble fractions were generated as described in Chapter 3) and N1, N2 and N3 fragments as GST fusions (GST-N1, GST-N2 and GST-N3). Soluble fractions of cell lysates having Strep fusions (100 µl) were incubated on the plate for 16 hours at 4 ºC followed by washing of the plate thrice with PBS-T. This was followed by incubation of soluble fractions from the cell lysate carrying GST fusion protein partners (100 µl) for 2 hours at 4 ºC. After washing three times with PBS-T, the plates were incubated with 100 µl mouse anti-GST monoclonal antibody (1:10,000; Sigma Aldrich) followed by 100 µl anti mouse HRP conjugated secondary antibody (1:2000; G biosciences). The plate was washed thrice with PBS-T followed by three washes with 1X PBS. The reaction was developed by addition of TMB as substrate for HRP and finally the reaction was stopped by 2N HCL and absorbance measured at 450nm. Binding of only Strep (Strep-N, Strep-P, Strep-M and Strep-G) and only GST (GST-N1, GST-N2 and GST-N3) fusions directly to the Streptactin plate was analysed as control. The test interaction pairs tested and checked for association in Chapter 5 (NP, PP) and
the non-interacting pairs (PM, PG) as well were included in the assay as controls (Strep-N+GST-P, Strep-P+GST-P, Strep-M+GST-P and Strep-G+GST-P). The binding of N1 region with a non-relevant protein [nonstructural protein 1 (nsP1) of Chikungunya virus (CHIKV) as Strep fusion] was also taken to observe the specificity of the interactions.
6.3 Results

6.3.1 Model building and validation

Structures for all four CHPV proteins were generated using VSV protein structures as templates. CHPV P protein is 293 aa in length and constitutes 5 cysteine residues, of which DiANNA server predicted 4 to be involved in intramolecular disulfide bond formation. In concordance with the predicted disulfide linkages, CHPV P protein structure was transformed to introduce disulfide links between Cys37-Cys172 and Cys57-Cys286. The analysis of the structure by Ramachandran plot displayed approximately 98% of the residues in allowed region (favored + allowed regions) and the rest as outliers. This could be expected as 52% of the structure is predicted to be loop, with almost all the outliers belonging to the loop region. Loop regions are the most variable regions and the outliers are common in these regions of proteins, however, the percentage of outliers can be reduced by refining the loops. Approximately 99.3% residues of CHPV N protein lie in the allowed regions. The structure analysis by Verify3D and ERRAT indicated all the protein models (CHPV N, P, M and G) to be of good quality with minimal inter-atomic clashes. The consecutive CHARMM based minimization of the protein models resulted in further structural stability and improved the physical realism and structural accuracy.

The structural model of CHPV N protein (Figure 6.1, i) helped in deciding the boundaries for the division of N gene into fragments on the basis of PiSQRD results. The regions were separated and all the side-chains of the isolated fragments were modified based on SEQRES information (Figure 6.2) using Accelrys Discovery Studio. The structures of the N fragments [N1 (Nt arm), N2 (Nt lobe) and N3 (Ct lobe)] were also generated as individual units. N1 fragment consists of a pair of anti-parallel beta strands, N2 is composed of 13 helices, one pair of parallel beta sheets and one anti-parallel beta sheet pair and N3 is formed by 10 helices (Figure 6.3 i, ii and iii). CHPV and VSV N protein structures were significantly similar as displayed by rigid structural alignment with RMSD of 1.77 over 422 amino acid residues.

6.3.2 Rigid-body docking

Structures of CHPV N, P, M and G proteins generated by homology modeling [N (Figure 6.1, i), P (Figure 6.1, ii), M (Figure 6.1, iii) and G (Figure 6.1, iv)] were docked on CHPV N protein model using ZDOCK at a 15° rotational sampling density. Out of 3600 generated conformations, top 2000 poses were further re-ranked (ZRank) using detailed
electrostatics, van der Waals, and desolvation energy terms. The success of the resulting predictions was evaluated based on their RMSD values. When the RMSD of one of the proteins is ≤ 10Å from the cluster center, the pose was defined as acceptable (the choice of 10Å RMSD is in concordance with the ligand RMSD used by CAPRI to define an acceptable solution in protein-protein docking). The pose generated after RDOCK refinement with minimum E_RDOCK score, was chosen as near-native structure for each interaction (Figure 6.4). RDOCK scores of all the complexes predicted a strong binding between the interacting pairs. A lower RDOCK score indicates more stable association of the proteins in a complex. Each of the top-ranked poses was found to be associated with large size clusters with very favorable electrostatic and atomic contact energy scores. Extensive protein side-chain movements among the binding interface residues were observed in all the protein complexes after the RDOCK refinement, while the backbone conformation remained largely unchanged. It was observed that N terminal fragment (Nt arm) of N protein (N1) interacts with N, P and M proteins while Nt lobe (N2) interacts with N, P and G proteins and Ct lobe (N3) interacts with N, P, M and G proteins. These interactions were experimentally checked by Y2H and ELISA methods.

![Figure 6.1: Structures of CHPV N, P, M and G proteins predicted using SWISS-MODEL workspace](image)

(i) Nucleoprotein, (ii) Phosphoprotein, (iii) Matrix protein and (iv) Glycoprotein of Chandipura virus rendered in cartoon (rainbow colour) form. N and C terminals are indicated.
**Figure 6.2:** Schematic representations of CHPV N1, N2 and N3 fragments

N terminal arm (N1; in red) is of 30 amino acids (aa). N terminal lobe (N2; in green) is of 278 aa. C terminal lobe (N3; in blue) constitutes the C terminal 193 aa of the CHPV N protein. The N terminal 10 aa residues and the C terminal 68 aa residues of N2 overlap with N1 and N3, respectively.

**Figure 6.3:** Structures of N1, N2 and N3 fragments of CHPV N protein using SWISS-MODELworkspace

(i) Nucleocapsid fragment 1 (Nt arm) (ii) Nucleocapsid fragment 2 (Nt lobe) and (iii) Nucleocapsid fragment 3 (Ct lobe) rendered in cartoon (rainbow colour) form.
Figure 6.4: Docking results for NN, NP, NM and NG interactions using ZDOCK and RDOCK
(a) Top rank NN complex ZDOCK pose 10 (E_RDOCK score -29.6347948). (b) Top rank NP complex ZDOCK pose 220 (E_RDOCK score -27.93216). (c) Top rank NM complex ZDOCK pose 684 (E_RDOCK score -31.35601). (d) Top rank NG complex ZDOCK pose 5 (E_RDOCK score -24.12947). CHPV proteins are represented as cartoon models. CHPV N protein is shown in blue and another CHPV interacting protein is shown in red. The interacting residues in the NN residual interface region are highlighted in boxes. Side chains of the N protein amino acids contributing to hydrogen bond formation with the residues of other protein are represented as a stick model (rendered in green). (a) The residues of CHPV N protein in the NN interface correspond to N1, N2 and N3 regions (b) The residues of CHPV N protein in the NP interface correspond to N1, N2 and N3 regions (c) The residues of CHPV N protein in the NM interface correspond to N1 and N3 regions and (d) The residues of CHPV N protein in the NG interface correspond to N2 and N3 regions.
6.3.3 Generation of nucleocapsid fragment fusions for Y2H and ELISA

The nucleocapsid fragments (N1, N2 and N3) were generated as BD, AD and GST fusions for carrying out the Y2H screening and ELISA assay. The three overlapping regions were PCR amplified with specific primers (Table 6.1) for each vector i.e., pGBKTK7 and pGADT7 of Y2H system and pGEX-4T3 bacterial expression vector. The N1 (90 bp), N2 (834 bp) and N3 (579 bp) gene fragments were observed at their expected sizes after PCR amplification on 1.2% agarose gel (Figure 6.5 and 6.6). The PCR products were purified and digested with Nde I and BamHI and cloned in pGBKTK7 and pGADT7 vectors to generate recombinants BD-N1, BD-N2 and BD-N3 for pGBKTK7 and AD-N1, AD-N2 and AD-N3 for pGADT7. For cloning in bacterial expression vector, PCR products were digested with BamHI and Xho I and cloned in pGEX-4T3 vector to generate GST-N1, GST-N2 and GST-N3, respectively. The recombinant clones for N1, N2 and N3 fragments in BD vector were screened by digestion of plasmid DNA isolated from transformed colonies with BamHI and Xho I. The released fragment of the respective gene size with additional 330 bp derived from plasmid backbone confirmed the cloning. N1 fragment was observed at 420 bp (Figure 6.7 lane 4-7), N2 at 1.1 kb (Figure 6.7 lane 1-3) and N3 at 909 bp (Figure 6.8 lane 1-5). Cloning of these fragments in AD vector was however confirmed with Hind III enzyme which releases a DNA fragment of 800 bp from the control vector and the fallout released for the N gene fragments were 800 bp in addition to the respective gene fragment sizes i.e., 890 bp for N1 (Figure 6.9 lane 10), 1.6 kb for N2 (Figure 6.9 lane 7, 8) and 1.3 kb for N3 (Figure 6.9 lane 2, 3 and 4), respectively. The clones generated in pGEX-4T3 vector, i.e., GST-N1, GST-N2 and GST-N3 were screened by digesting the plasmid minipreps with BamHI and Xho I. The digestion by these enzymes released a segment of the respective gene size N1 (90 bp), N2 (579 bp) and N3 (834 bp) (Figure 6.10 lane 3, 4 for N1, lane 5, 6 for N2 and lane 7, 8 for N3 fragment). GST-N1, GST-N2 and GST-N3 were checked for the protein expression by inducing the E. coli BL21 cells harbouring them with 1 mM IPTG. The expressed proteins were checked for their solubilisation after cell lysis and all of them were found to be partially soluble i.e., were present in the supernatant as well as cell pellet fractions (Figure 6.11 a, b and c). The soluble fractions were used for ELISA assay.
N1, N2 and N3 fragments of N gene were PCR amplified with specifically designed primers for cloning in BD and AD vectors. All the three N gene fragments were observed at expected sizes; N1 (90 bp), N2 (834 bp) and N3 (579 bp). L1 is 1 kb DNA ladder (Fermentas; Molecular sizes are indicated).

PCR amplified N1, N2 and N3 gene fragments with specific primers for cloning in pGEX-4T3 bacterial expression vector. All the gene fragments were observed at expected sizes; N1 (90 bp), N2 (834 bp) and N3 (579 bp). L1 is 1kb DNA ladder and L is 100 bp DNA ladder (Fermentas; Molecular sizes are indicated).
Figure 6.7: Restriction digestion of BD-N1 and BD-N2 using BamH I and Xho I
Recombinants for BD-N1 and BD-N2 were screened by restriction enzyme digestion using BamH I and Xho I. Lanes 1-3 represent the miniprep DNAs digested to screen BD-N2 clones and show the released fragment at 834 bp (gene size) with additional 330 bp (1.16 kb). Lanes 4-8 represent the plasmid DNA screened for BD-N1 clones and lanes 4-7 show the released fragment at 90 bp (gene size) with additional 330 bp from vector backbone (0.42 kb). L1 is 1 kb DNA ladder (Fermentas; Molecular sizes are indicated). Lane 10 is the control BD vector digested with the same enzyme combination.

Figure 6.8: Restriction digestion of BD-N3 using BamH I and Xho I
The plasmid DNAs were prepared from the transformed colonies and screened for recombinants of BD-N3 by restriction enzyme digestion using BamH I and Xho I. Lanes 1-5 represent the recombinants for BD-N3 showing the released fragment at the fragment size (579 bp) with additional 330 bp from vector backbone (0.9 kb). L1 is 1 kb DNA ladder (Fermentas; Molecular sizes are indicated). Lane 6 is the control vector.
Recombinants for AD-N1, N2 and N3 were screened by restriction enzyme digestion of the plasmid DNAs prepared from the transformed colonies using Hind III. The released fragments were observed at the respective sizes with additional 800 bp from vector backbone. Lanes 1-4 are the minis tested for AD-N3 recombinants (1.37 kb) while lanes 5 and 8 are for AD-N2 clones (1.6 kb) and lanes 9-11 are for AD-N1 (0.9 kb). L1 is 1 kb DNA ladder (Fermentas; Molecular sizes are indicated). C is the control AD vector digested with Hind III.

Recombinants for GST-N1, N2 and N3 were screened by restriction digestion using BamH I and Xho I. Lanes 3 and 4 are the minis tested for GST-N1 clones while 5 and 6 are for GST-N2 and lanes 7 and 8 are for GST-N3. Lane 2 represents the empty GST vector as control. The released fragment was observed at the respective gene sizes. L1 is 1 kb DNA ladder and L is 100 bp DNA ladder (Fermentas; Molecular sizes are indicated).
E. coli BL21 cells harbouring GST-N1, GST-N2 and GST-N3 fusion plasmids were induced with 1mM IPTG. Cells expressing these fragments were lysed to check for the solubility of proteins. Before induction (BI) and after induction (AI) samples and the cell pellet (Pl) and supernatant (Sup) samples obtained after cell lysis were analysed by western blotting and the proteins were detected with anti GST antibody. (a) Induction and solubilisation analysis of N1, (b) N2 and (c) N3 proteins, respectively. Prestained protein ladder (Fermentas) was used in all blots and the ladder bands are indicated in kDa.

6.3.4 Y2H analysis - transformation, mating and interaction screening

The recombinants BD-N1, BD-N2 and BD-N3 derived from pGBKT7 and AD-N1, AD-N2 and AD-N3 from pGADT7 were transformed in Y187 and AH109 yeast cells, respectively. The transformants were plated on SD/-Trp media for BD fusions and on SD/-Leu media for AD fusions. Before proceeding for the Y2H interaction analysis, BD and AD fusion constructs were checked for their ability to activate the reporter gene expression (autoactivation) on SD/-Trp/-His and SD/-Leu/-His, respectively. None of the recombinant BD or AD vectors tested in the study activated reporter gene and showed background transcriptional activity and thus used for Y2H studies. Full length CHPV viral genes had been checked for autoactivation in Chapter 4. BD-P was found to activate the histidine reporter
gene hence combinations in the reverse direction i.e., involving those of AD-P were considered.

Y187/AH109 cells carrying bait and prey recombinants, respectively, were allowed to mate and the resulting diploids were screened for their ability to grow on SD media lacking tryptophan and leucine (SD/-Trp/-Leu) [Figure 6.12]. The 12 unique pairwise interactions of N1, N2 and N3 with full length N, P, M and G proteins were tested from both directions i.e., each protein was taken as bait as well as prey fusion. Mated diploids were checked for putative positive interaction by analyzing expression of reporter genes HIS3 and MEL1. The expression of HIS3 was checked by growth of diploid cells on SD/-Trp/-Leu/-His media (Figure 6.13). The colonies which grew on histidine deficient media were considered to be expressing interacting proteins. Another reporter gene i.e., MEL1 was tested for activation by plating on SD/-Trp/-Leu/-His/α-gal media and blue colored colonies indicated positive interactions (Figure 6.14). It was observed that AD-N1 and AD-N2 were interacting with empty bait vectors and were able to switch on the reporter gene (Figure 6.14 sector 28 and 29). Thus the BD interaction pairs of AD-N1 and AD-N2 showed interaction on the selection media (Figure 6.14 sector 13-20). The results generated using N1-AD, N2-AD and those involving BD-P due to autoactivation were not considered. The control vectors for positive interaction (pGBK7-53 and pGAD7T-T) encoding tumor suppressor protein p53 and simian virus large T-antigen (Figure 6.14, sector 31) and the control vectors for negative interaction (pGBK7-Lam and pGAD7T-T) encoding human lamin C and Simian virus 40 large T-antigen (Figure 6.14, sector 32) gave positive and negative reactions, respectively as expected.
Figure 6.12: Y2H mating results of N1, N2 and N3 with full length CHPV N, P, M and G proteins. All the possible combinations of Y187/AH109 cells expressing N1, N2 and N3 were mated with AH109/Y187 cells expressing full length CHPV N, P, M and G proteins and streaked on SD/-Trp/-Leu plates for diploids selection. Panel I (a), (b) and (c) are BD-N1, BD-N2 and BD-N3 mated with N, P, M and G proteins as AD fusions respectively. Panel II (a), (b) and (c) are AD-N1, AD-N2 and AD-N3 mated with N, P, M and G proteins as BD fusions.
Figure 6.13: CHPV-N fragments mapping by Y2H

All four viral genes (N, P, M and G) cloned with AD (prey) and transformed in AH109 yeast strain were mated with BD-N1, BD-N2 and BD-N3 and transformed in Y187 yeast strain. The interactions were analyzed on SD-/Trp-/Leu-/His media. The table indicates the scheme of various pairs in the different panels. Presence of growth on medium (sectors 1, 2, 3, 4, 5, 8, 9, 10 and 11) indicated interaction between the proteins while absence (sectors 6, 7 and 12) suggested no interaction.
Figure 6.14: X-alpha galactosidase assay for interaction confirmation

All possible interacting pairs of CHPV nucleocapsid protein with other viral proteins with controls were plated in an array format on plate containing X α-gal (SD/-Trp/-Leu/-His/α-gal). Interactions were considered in both directions i.e., as BD and AD fusions. The interaction of N protein fragments with empty BD and AD vectors and the interaction of pGBK7-53+pGADT7-T and pGBK7-lam+pGADT7-T were included in the assay as controls.
6.3.5 Interaction screening by ELISA

ELISA was performed to identify the regions of N protein responsible for its interactions with other four viral proteins. In this assay, N, P, M and G proteins as Strep fusions and N fragments as GST fusion were utilised. The cell lysates having soluble Strep-N, Strep-P, Strep-M and Strep-G were allowed to bind to Streptactin coated microtiter plates. This was followed by incubation with cell lysates having soluble GST-N1, GST-N2 and GST-N3. Binding of only Strep and only GST fusions to the Streptactin coated plate served as controls. The interactions were checked using mouse anti GST monoclonal antibody (Figures 6.15 and 6.16).

A total of 9 protein interactions were identified after analysis of 12 unique putative protein interaction pairs. All the pairs considered in Y2H screening were checked independently by ELISA. The overall interaction data obtained from ZDOCK/RDOCK, Y2H analysis, ELISA assay and literature has been assembled in Table 6.2. The data from Y2H and ELISA experiments corroborated well with each other with no exceptions. The interactions observed by these two independent techniques were also found to be in concordance with the bioinformatics based predictions (75% accuracy) and the available literature on the virus (100%). Interestingly, the N1 fragment was shown to interact with all four viral proteins i.e., N, P, M and G. A non-relevant protein (nsP1 as strep fusion) of Chikungunya virus was considered in ELISA to rule out the possibility of non-specific binding of N1 fragment. CHPV N2 fragment was shown to bind N and G proteins while N3 associated with N, P and M proteins (Table 6.2).
**Figure 6.15**: Validation of N protein fragments interactions with CHPV full length proteins by ELISA.

The lysates of Strep tagged viral proteins (N, P, M and G) were incubated onto Streptactin coated microtiter plate followed by incubation with GST fusion N protein fragments (N1, N2 and N3). The interacting GST fusion protein was detected using anti GST monoclonal antibody. The lysates of only Strep (without any GST fusion protein lysate; B1, C1, D1 and E1) and only GST fusions (without any Strep fusion protein lysate; A2, A3 and A4) were taken as controls. Blue colour production after addition of substrate (TMB) indicated a positive interaction (Panel I). A1 was air blank. Panel II shows the final product after stopping the reaction with HCl in corresponding lanes.
**Figure 6.16**: Graphical representation of ELISA for validation of CHPV N protein fragments interactions with full length proteins

ELISA was performed to validate the interactions between three N protein fragments (N1, N2 and N3) and full length CHPV N, P, M and G proteins. Only Strep fusion proteins (Strep-N, Strep-P, Strep-M, Strep-G and Strep-nsP1) and only GST fusion proteins (GST-N1, GST-N2, GST-N3 and GST-P) were considered as controls. A non relevant strep tag fusion protein (Strep-nsP1) of Chikungunya virus was taken to test the specificity of N1 interactions. The data shown is mean of three independent experiments (mean ± SD). The known interaction pairs (PN and PP) and non interactors (PM and PG) of CHPV were included as experimental controls. The absorbance at 450 nm is plotted on the Y axis (10 division = 0.2 OD) and the protein pairs are considered on the X axis.
Table 6.2: Cumulative results of CHPV N protein fragments interaction analysis

<table>
<thead>
<tr>
<th>Protein pair used for interaction analysis</th>
<th>Interaction analysis by ZDOCK and RDOCK</th>
<th>Y2H assay</th>
<th>ELISA</th>
<th>Known from Literature</th>
</tr>
</thead>
<tbody>
<tr>
<td>N1-N</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+ (Mondal et. al., 2016)</td>
</tr>
<tr>
<td>N1-P</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+ (Mondal et. al., 2012)</td>
</tr>
<tr>
<td>N1-M</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N1-G</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>N2-N</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+ (Mondal et. al., 2010)</td>
</tr>
<tr>
<td>N2-P</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>N2-M</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>N2-G</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>N3-N</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+ (Mondal et. al., 2016)</td>
</tr>
<tr>
<td>N3-P</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+ (Mondal et. al., 2012)</td>
</tr>
<tr>
<td>N3-M</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>N3-G</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

+: represents positive interaction

-: represents negative interaction
6.4 Discussion

During the infection and multiplication of viruses, the proteins of viruses extensively interact with one another to perform their functions. These protein-protein interactions (PPIs) occur as a result of physical interactions mediated through particular domains present in the proteins. For an interaction network, the domains which mediate different interactions could be considered as building blocks. For several viruses domain mapping has been carried out, examples include Herpes Simplex Virus type I (HSV), Epstein Barr Virus (EBV), Kaposi’s sarcoma associated Human Virus (KSHV) [176], Murine Coronavirus [177], Rabies Virus (RV) [165], VSV [40, 81], Sendai virus [178]. The mapping of domains has been performed for N and P proteins of CHPV as well [19, 88]. These studies have highlighted the importance of identifying the regions or domains which can be targeted for therapeutic strategies.

The aim of this chapter was to extend the work on the observation that the CHPV N protein interacts with four of the viral proteins i.e., N, P, M and G through domain mapping studies. Different approaches for interaction studies like bioinformatics based docking studies, yeast two-hybrid system and ELISA were employed which helped in the identification of specific regions of N protein involved in mediating its interactions with other CHPV viral proteins. Moreover, the structures of CHPV proteins predicted in the present study using computational methods provided the first insight into the less known structures of CHPV proteins. The analysis of the interactions among N protein regions and CHPV N, P, M and G proteins using three different and independent techniques (Bioinformatics predictions, Y2H and ELISA) helped to assess the overlap and hence reliability of data. The correlation of the docking based predictions with the interaction data from Y2H and ELISA was observed to be 75% which in turn matched 100% with the known literature (Table 6.2).

The study identified the interacting residues involved in NN association which are present in all the three fragments of nucleocapsid protein considered in this study (N1, N2 and N3). However, the central 278 aa region (N2) essential for interaction with G protein is shown to be dispensable for interactions with M and P proteins.

Mapping studies of CHPV N protein have been carried out earlier at the domain level, however, they involved the generation of N domains by random enzymatic digestion [19, 88]. On the other hand, the division of nucleocapsid protein into N1, N2 and N3 regions was based on precise structural and biophysical aspects. The interacting residues involved in NN association identified in the study lie within the N terminal and central regions of the N monomer and similar results were shown by Mondal and co-workers [19]. The bioinformatics
based predictions in the present study have further narrowed down these regions to smaller peptides including residues 8 to 22 at the N terminus and 245 to 256 in the central region. Moreover, with evidence from the oligomerisation studies of VSV N protein [40], it is suggested in this study the involvement of intermittent residues from 321 to 395 at the C terminus in the oligomerisation of N protein. According to deletion studies by Mondal and coworkers the N terminal 1 to 180 aa and C terminal 320 to 390 residues of N protein are involved in NP interaction in CHPV and in this mapping study, on the basis of computational studies it is suggested that the smaller peptides within these regions – residues 2 to 30, 140 to 165, 205 to 240 and 320 to 343, are indispensable for this association as N protein oligomerisation as well as RNA binding constraints have also been considered for NP association. The data generated here corroborates well with the previously identified interacting domains involved in NN and NP interactions for both CHPV [19, 88] and VSV [40] thus validating the approach of domain interaction analysis adopted in the study. However, in addition to the identification of interacting regions for NN and NP associations, the regions responsible for NM and NG interactions have also been predicted. The NM interaction involves the N terminal 16 to 20 aa and C terminal 318 to 420 residues, while NG binding requires the 144 to 240 aa in the central region.

Although important data has been generated by mapping studies, the biological significance of these interactions is the scope of further experimentation. Nevertheless, these associations can prove to be valuable starting points for understanding CHPV biology and designing antiviral strategies. Components blocking the N protein interacting regions may represent a novel class of molecules suitable for a therapeutic intervention in Chandipura mediated disease. In this regard, revelation of a small N terminal fragment of 30 aa was important which was found essential in all the intraviral protein interactions involving N protein and might have the potential as a candidate antiviral target.