CHAPTER-2
LITERATURE REVIEW
2.1 Introduction

Cellular life forms possess the basic mechanism of genome replication and expression based upon the presence of double stranded (ds) DNA genome and positive strand messenger RNA (mRNA). In contrary to these life forms, viruses are broad and genetically distinct biological units and utilize all kinds of genetic material [positive-strand, negative-strand and double stranded (ds) RNA, single-stranded (ss) DNA, (ds) DNA] and execute all possible genetic methods (RNA or DNA replication and transcription, as well as reverse transcription of RNA to DNA). These various types of virus genomes are divided into seven fundamentally different groups according to Baltimore classification system, and they apparently require different mechanisms for their replication. The groups include (I) Double-stranded DNA viruses (e.g., Adenoviruses), (II) Single-stranded (+) sense DNA viruses (e.g., Paroviruses), (III) Double-stranded RNA viruses (e.g., Reoviruses), (IV) Single-stranded (+) sense RNA viruses (e.g., Togaviruses), (V) Single-stranded (-) sense RNA viruses (e.g., Rhabdoviruses), (VI) Single-stranded (+) sense RNA viruses with DNA intermediate in life cycle (e.g., Retroviruses) and (VII) Double stranded DNA viruses with RNA intermediate (e.g., Hepadnaviruses). In particular, viruses with RNA genomes that do not have a DNA intermediate during their life cycle are the simplest genetic elements and the majority of human and animal pathogenic viruses have RNA genomes. Some of the greater known examples of diseases caused by these human pathogens include influenza, measles, mumps, rubella, polio, rabies, yellow fever, dengue and ebola hemorrhagic fever. Together, these viruses have been responsible for millions of mortalities worldwide. As a matter of more concern, most of the emerging viruses are also RNA viruses.

Among the RNA viruses, negative strand RNA viruses are distinguished for their ability to undergo frequent mutations which is the underline factor for great diversification in this group [22]. An important determinant for this genetic variability is the error prone nature of the RNA-dependent RNA polymerase, which may result in high rates of mutation, sequence deletion, insertion, and recombination between genomes during virus replication. Due to the absence of efficient proofreading and post-replication repair activities associated with RNA replicases, mutation rates of RNA viruses have been estimated to be very high. As a result, viral populations are constantly changing and the ensuing genetic diversification is being driven by evolution. In the long term, this leads to the segregation of new variants, species or genus. The first order to be designated in the virus taxonomy was order Mononegavirales and it represents the entire group of negative sense single strand RNA
viruses [23]. Rapid evolution and accordingly high diversity are the major driving forces of virus virulence, tropism, host range and transmission, and thus the perpetuation of the mononegavirales.

The order Mononegavirales comprises four families namely Rhabdoviridae, Paramyxoviridae, Filoviridae and Bornaviridae. Rhabdoviridae is the largest viral family and includes a diverse range of viruses which can infect a very wide host range including plants, animals, fishes and humans. Rhabdoviridae includes following genera; Vesiculovirus, Lyssavirus, Ephemerovirus, Novirhabdovirus, Cytorhabdovirus, Nucleorhabdovirus and Dichorhabdovirus. Rhabdoviruses, like other mononegavirales, are enveloped negative ss (single stranded) RNA viruses with a lipid bilayer envelope derived from the plasma membrane of the infected host cell. The virus particles are bullet-shaped particles (characteristic feature of the family) measuring approximately 100–430 nm in length and 45–100 nm in diameter. The most studied viruses of this family are the prototypes Vesicular Stomatitis Virus (VSV) [Genus Vesiculovirus] and Rabies virus (RV) [Genus Lyssavirus]. Chandipura virus (CHPV), isolated in 1965 and grouped with rhabdoviruses in the Vesiculovirus genus [6, 7], had not gained much attention until the outbreak of 2003 in the central states of India, during which the virus was associated with human encephalitis for the first time [4].
2.2 Chandipura virus

Chandipura virus has been placed in genus *Vesiculovirus* (Figure 2.1) on account of its similarity with the prototype virus of the genus i.e., Vesicular Stomatitis Virus (VSV) based on protein and nucleic acid composition.

![Figure 2.1: Classification of CHPV](image)

2.2.1 An epidemiological outlook

Chandipura virus is a recently emerged human pathogen associated with a number of encephalitic outbreaks in different parts of India [4, 5, 24]. The virus was first isolated in 1965 from two adult patients suffering from febrile illness during an outbreak of Chikungunya and Dengue viruses in Chandipura village of Maharashtra state, India [6]. CHPV was later isolated from sandflies in 1967 from Aurangabad district [25] followed by its isolation from serum sample of a child suffering from encephalopathy in Madhya Pradesh [26]. The first evidence for its association in human epidemics was observed almost four decades later in 2003, when CHPV was identified from patient samples during an outbreak of acute encephalitis with fatality rate of 55.6% in Andhra Pradesh [4]. Subsequently, another outbreak of CHPV infection with 44% fatality rate was reported in Maharashtra [4]. In 2004, CHPV outbreak was reported in the eastern districts of Gujarat state with a case fatality rate of 78.3% [5]. In the following years, Chandipura virus mediated encephalitis cases were reported in the states of Andhra Pradesh and Maharashtra [24]. The most recent case has been reported in 2012 where the virus has claimed 21 lives of children in Maharashtra state with a fatality rate of
58% (unpublished data). These occurrences indicate the emergence of Chandipura virus as a lethal virus in Indian subcontinent. Rapid onset of disease i.e., within 24-48 hours, high fatality rate and age dependent pathogenesis have been found to be the key features of CHPV mediated encephalitis.

2.2.2 Chandipura viral infection

Age dependent susceptibility has been observed in all the outbreaks of CHPV [4, 5, 24]. Although, IgM antibodies to CHPV were detected in the blood of adults in endemic areas, no encephalitis cases have been reported to date [4, 5]. The encephalitis cases occurred only among children of age ranging from 9 months to 15 years and the majority of deaths occurred within 48 to 72 hours of hospitalisation in all the outbreaks [4, 5, 24].

The exact route of pathogenesis by which the virus reaches the brain is still unknown. Although the virus replication in the brain has been found to be responsible for neurological symptoms and mortality [27]. VSV, the virus closest to CHPV, takes the olfactory route to reach the brain and Rabies Virus, the only other human pathogen in the Rhabdoviridae family, enters the motor neurons from the muscles and travels in a retrograde manner to reach the brain. To have an insight into the viral pathogenesis, studies were conducted on murine models and it revealed that young mice were susceptible to CHPV through intravenous, intraperitoneal and intracerebral routes of infection, while the adult mice were susceptible only to the intracerebral route of infection [27]. In young mice, the intravenous route of infection produced viremia and the virus was observed to cross the blood brain barrier to replicate in the central nervous system [27]. It was observed that maximum viral titer reached in the blood 24 hours post infection. Circulating viruses were observed to be effectively cleared by virus specific IgM antibody and by 48 hours the viral titer values in the blood dropped with a corresponding increase in the brain. Within 72 hours most of the mice were dead due to encephalitis and the autopsy results revealed maximum viral load in the brain with comparatively low levels in the blood [27]. The short duration of pathogenesis (72 hours) and the high viral titer levels in the blood within 24 hours indicate circulatory route may be the path of central nervous system (CNS) invasion for CHPV. In-vitro studies have proven the ability of the virus to multiply in macrophages, T cells and B cells [27], but it is not clear whether the virus just uses circulatory system as a path to reach the brain or uses the blood cells as a means of transport along with using it as a local replication site before entering the CNS.
2.2.3 Potential vectors of Chandipura virus

Dhanda and co-workers described that CHPV is transmitted to humans by sandflies (\textit{Phlebotomus papatasi}) [25]. However, in 1967 Rao and coworkers had shown experimental transmission of CHPV by mosquitoes [28]. Later it was demonstrated in laboratory studies that \textit{Aedes aegypti} was the most efficient in transmitting the virus to mice than the \textit{Culex tritaeniorhynchus}, \textit{Culex Bitaeniorhynchus} and \textit{Culex quinquefasciatus} [29]. Further in 2005, it was shown that CHPV is transmitted in \textit{Aedes aegypti} mosquito by vertical and veneral transmission [30]. However, there is no evidence for the role of mosquitoes in CHPV mediated disease transmission in field.

The geographical isolation of the vector (sandflies) along with optimum climatic conditions might be the reason for CHPV to be endemic to India particularly in the states along the Tropic of Cancer. The sole reason for the absence of CHPV outbreaks in the European countries is the absence of the vector in temperate climates. Although CHPV has been isolated from sand flies in the South African region, encephalitic outbreaks have only been reported in the Indian subcontinent. These insects are prevalent in India in the late summer and early monsoon period, a season that coincides with the outbreaks of CHPV mediated disease.

2.2.4 Viral Structure and Genome Organization

Chandipura virus has a typical bullet shaped structure, 150-165 nm long and 50-65 nm wide, bound by a lipoproteinaceous envelope. The envelope has distinct surface projections of 9-11 nm in size which are trimers of glycoprotein and protrude externally. Matrix protein lies on the inner bilayer of the membrane. Enclosed inside the envelope is a helical ribonucleoparticle (RNP) which is a non-segmented single strand negative sense RNA enwrapped by Nucleocapsid protein. Large protein and Phosphoprotein are packaged within the mature virion and remain associated with the core nucleocapsid particle (Figure 2.2 and 2.3).
Figure 2.2: Schematic representation of a typical bullet shaped Chandipura virus

Figure 2.3: Transmission electron micrographs of CHPV isolates from tissue culture
A: Two negatively stained CHPV particles. B: Negative stained CHPV particle showing typical vesiculovirus morphology. Inset shows a virus particle with a released helical ribonucleoprotein coil [4]. The image was freely available as a Google image.

CHPV genome comprises of a negative sense single stranded RNA molecule of approximately 11 kb [8]. Five proteins, Nucleocapsid protein (N), Phosphoprotein (P), Matrix protein (M), Glycoprotein (G) and Large protein (L) are encoded from the transcription of viral genes in sequential manner from a single promoter at 3’ end of the genome resulting in the decreased amounts of each transcript in the order 3’- Le-N-P-M-G-L-5’ (Figure 2.4). Le is a 49 nucleotide leader RNA which is transcribed from the 3’ terminus of the genome and remains untranslated, uncapped and non-polyadenylated. Similarly a trailer (Tr) sequence, of 46 nucleotides, is present at the 5’ end of the genome which remains untranscribed.
CHPV has marked resemblance with the prototype virus of the genus, VSV, but can be distinguished on the basis of serology and its ability to infect humans. Interestingly, it has been revealed by comparative sequence analysis that CHPV is evolutionarily equidistant from the new world Vesiculoviruses i.e., VSV Indiana (VSVind) and VSV New Jersey (VSVnj) and is closely related to the Asian Vesiculovirus, Isfahan virus [8]. VSV remains one of the extensively studied viruses of the family and is being reviewed in following sections.

2.3 CHPV proteins: Functions

2.3.1 Nucleocapsid Protein (N)

N protein is the most abundant protein in the virus infected cells and accounts for the major protein mass (90%) of the purified RNP core [31]. The viral genomic RNA is encapsidated by the N protein in a stoichiometry of 9 nucleotides per N monomer to form a tight helical structure [31]. CHPV N gene was sequenced in 1987 and it showed ~ 50% sequence homology with VSVind and VSVnj serotypes [32]. The nucleocapsid protein tends to form aggregates and the aggregated protein does not encapsidate viral RNA which becomes prone to degradation by cellular RNases. The coexpression of N and P proteins of VSV resulted in encapsidation of viral RNA [33] indicating the importance of P protein in maintaining the functional stability of N protein (Figure 2.5). CHPV N gene was cloned in 2001 and the biophysical and biochemical properties were studied in detail [20]. It is a 422 amino acid protein with a molecular weight of 47.9 kDa and the major biological function of N protein is to enwrap the viral RNA and protect it from degradation by cellular RNases. The encapsidation of replication product by N protein is concurrent with genomic RNA synthesis.
forming a precise structure [34, 35, 36] which is compared to histone mediated enwrapping of DNA molecule [7, 37].

The N protein plays a dual role by its ability to recognize specific sequence on nascent RNA, a process termed as nucleation and progressively encapsidate RNA, a process termed as elongation. N protein in its monomeric state recognizes a specific sequence within the first 21 nucleotides of the leader RNA which is not recognized by the oligomerised N protein [38]. During the nucleation step, N monomer recognizes specific sequences on nascent viral leader RNA to initiate nucleocapsid assembly. During elongation phase the N-N association results in both inter and intra molecular conformational changes that enable the newly polymerized N protein to bind to the heterogeneous sequences on the RNA molecule. While the continual supply of N monomers is guaranteed by N-P complex, the N-N as well as the N protein associated with RNA (N-RNA), traps N from N-P complex within encapsidation complex during elongation step [3].

![Figure 2.5: Self assembly of N protein monomer into oligomer that binds to RNA](image)

A. N protein molecules formed after translation.
B. 1. P protein binds to N monomer and prevents its self aggregation.
C. Initiation of the encapsidation process (Nucleation).
D. Constant supply of N protein in monomeric form by P protein continues the process of encapsidation.

The N-P and N-RNA interactions in VSV [39, 40], RV [39] and CHPV [3, 19] have been demonstrated. The domain-domain interactions involved in N protein oligomerization (N-N interaction) have been elucidated for VSV [40]. In N protein, the contacts involved in the domain-domain interactions are divided into three groups (contact I, II and III) that span among four neighboring N monomers in a repeated fashion [40]. The contact I is between the
N terminal arm of monomer x with the C terminal lobe of the monomer to the left (x-1); contact II is between the extended C terminal loop of the monomer x with the C terminal lobe of the monomer to the right (x+1) and contact III is between the N terminal arm of the monomer x with the extended C terminal loop of the monomer x-2 [41]. Although the functional domains involved in N protein oligomerization of CHPV have been shown [19], the actual domain-domain interactions associated with the interactions of one N monomer molecule with four neighboring N monomers in the RNP helix of CHPV have not yet been studied. Mondal and co workers in 2010 had suggested that the N terminal region of the nucleocapsid protein containing the first 47 amino acids and a central region comprising residues 180-265 are critical for the self assembly of N protein [19].

2.3.2 Phosphoprotein (P)

The 32.5 kDa phosphoprotein (P) is the catalytic subunit of the viral RNA dependent RNA polymerase (RdRp). The phosphorylation state of the P protein determines the transcriptase and replicase activity of RdRp. P protein has also been shown to act as transcription-replication switch (Figure 2.6) as the protein in its phosphorylated multimeric state (P1) forms a complex with the L-protein to produce functional transcriptase [42], whereas in its unphosphorylated state (P0), it complexes with the L-protein to form replicase [43]. VSV P protein was found to be extensively phosphorylated at multiple sites and casein kinase II was responsible for phosphorylating P protein [37, 44]. The site which was phosphorylated by CK II was identified to be serine 62 at the N terminal domain in CHPV P protein [21, 45]. The P protein also interacts with N protein through its C-terminus and keeps it in a soluble and active form capable of encapsidating viral RNA. In CHPV, P protein acts like a chaperone and plays a key role in the folding of nucleocapsid protein [20].

P protein by virtue of its interactions with N and L proteins plays an important regulatory role in the life cycle of vesiculoviruses. In a recent work by Kumar and Arankalle, CHPV P protein was selected as the target for siRNA based gene silencing strategy to fight against chandipura virus encephalitis. P protein selection was based upon its indespensable role in the activity of viral RdRp [46].

11
2.3.3 Matrix Protein (M)

The M protein (26.6 kDa) of VSV is known to play a critical role in the assembly as well as budding of virions from host cell membrane [13, 14, 47]. In assembled virions, M protein condenses the viral nucleocapsid and helps in its packaging. At the time of viral assembly, M protein binds with the RNP to condense into a tightly packed helix that results in the characteristic bullet shaped morphology of the virus [48]. Apart from functioning as a structural component, M protein also controls viral transcription by releasing the RNP core in the host cell [49, 50]. The concentration of M protein and RNP core are in equilibrium with one another. It has been observed that a high concentration of M protein inhibits viral transcription by favoring condensation rather than dissociation. They also regulate host transcription [16] and translation [15]; and hence are responsible for viral cytopathic effects. Interaction of VSV M protein with tubulin is responsible for the disruption of the host cytoskeleton, which results in rounding of infected cells [51]. VSV M protein can also block the mRNA export by disrupting function of Rae1/mrnp41 [52]. This protein interacts with eukaryotic Initiation Factor-3h (eIF-3h) subunit at various levels of translation, thus inhibiting host macromolecular synthesis. The M protein is essential in regulating the translation of host mRNA. This it does so by binding with the phosphorylation site of translation initiation factors and regulating them. Due to multiple roles that rhabdoviral M proteins play in the replication and pathogenesis of the viruses, it is speculated that CHPV-M may be involved in interactions with host proteins.
2.3.4 Glycoprotein (G)

Glycoprotein (69 kDa) is the sole spike protein that protrudes out from the viral membrane. G protein of VSV has been shown to play an integral role in virus adsorption, release of core particle and assembly and budding of virion particles. Both receptor identification and membrane fusion are mediated by this transmembrane protein [53]. Until recent years only two classes of viral fusion proteins had been described on the basis of their molecular architecture - class I (e.g., influenza virus hemagglutinin) and class II (e.g., E protein of flaviviruses and E1 of alphaviruses). The VSV G protein exhibits the properties of both the classes of viral membrane fusion proteins and is classified as class III viral membrane fusion protein. G protein consists of an N terminal signal peptide followed by three domains; ectodomain, a transmembrane domain and a cytoplasmic tail. The ectodomain of G protein has been further divided into: the lateral domain (domain I), the trimerisation domain (domain II), the PH domain (domain III), the fusion domain (domain IV) and the C terminal part (Cter) [Figure 2.7]. The crystal structure of the prefusion form of the VSV glycoprotein G (ectodomain) has been elucidated recently using electron microscope [53].

The G protein can reversibly adopt three different conformations; the native state present on the viral surface which is stable at pH 7, the activated state that fuses with the target membrane and a fusion inactive postfusion state that is stable at low pH conditions [53]. In the prefusion state, the G protein trimer assembled as spike resembles a tripod. Each leg of the tripod is composed of a fusion domain with the fusion loop pointing towards the viral membrane. In this tripod arrangement, the fusion domains are set wide apart, keeping the fusion loops separate from each other. The conformational changes involve remarkable reorganization of the G molecule. The post fusion conformation is achieved by flipping both the fusion domain and the C terminal region by 180˚ with respect to domain I and II. During the conformational change domain I, III and IV retain their folded structure [53, 54]. The N-terminal region of G protein has been shown to be crucial for membrane fusion activity and viral infectivity [55].

G protein is responsible for the production of neutralizing antibodies [56] and act as a major antigenic determinant [57, 58]. Recently, CHPV G protein has been expressed and purified using baculovirus expression system and is being tested as a candidate vaccine [59]. More recently, in another attempt for vaccine production, CHPV G protein has been purified using Pichia Pastoris expression system [60]. However these studies are still preliminary and require further experimental testing.
Figure 2.7: Structure of VSV (Indiana strain) G protein monomer

The figure shows the D I (domain I), D II (domain II), D III (domain III), fusion peptide (FP), D IV; (domain IV), C (C terminal region), N (N terminal region), FP (fusion peptide), Ex1(extension 1) is the linker between domain III and IV and Ex2 (extension 2) is the linker between domain II and III. The figure on the left depicts the prefusion native state and on the right depicts the postfusion state [61].

2.3.5 Large Protein (L)

L and P protein together constitute the viral transcriptase, L protein retaining the catalytic activity of RNA polymerisation, capping and polyadenylation. Unlike other four proteins of CHPV, L protein exhibits higher degree of homology with other rhabdoviruses. The conserved residues in VSV L protein in the central region [62] responsible for RNA polymerisation are also present in CHPV L protein [8]. Capping reaction catalyzed by L protein in VSV has evolved independent of eukaryotes. The RNA-dependent RNA polymerase L protein of VSV incorporates the GDP moiety of GTP into the cap structure of mRNAs instead of GMP as in eukaryotes [63]. A recent study carried out on CHPV L protein revealed that the L protein of CHPV exhibits a VSV-like RNA: GDP polyribonucleotidyltransferase (PRNTase) activity, which transfers the 5’P from viral mRNA start sequence to GDP to produce a capped RNA [12]. Another function of L protein is addition of poly (A) tail to the 3’ end of viral mRNA by polymerase slippage during transcription termination [10, 11]. VSV L protein is also shown to be associated with protein
kinase activity, whether intrinsic or due to some cellular kinase, termed as L associated kinase, LAK [7, 37].

VSV L protein also interacts with Hsp90, one of the host factors that is important for the replication of negative strand viruses. In VSV, Hsp90 stabilizes the L protein and thus has been targeted for antiviral studies. Hsp90 inhibitors like geldanamycin and radicicol decreased the half life of the L protein and it resulted in inhibition of viral replication by destabilizing viral L protein [64]. The functions of L protein of VSV have been demarcated by mutational studies to its six conserved domains. Domains II and III are implicated in RNA synthesis. Domain II binds RNA and domain III has GDNQ motif which catalyzes the formation of phosphodiester bonds. Domains I and V have been shown to link transcription and replication. Domain IV affects transcription termination whereas VI domain is required for cap methylation of viral mRNA [65].

2.4 Viral Life Cycle and Biology

The Vesiculovirus life cycle can be divided into eight distinct steps, namely adsorption of the virus particle, penetration of virus into the cell, uncoating and release of the core RNP into the cytosol, transcription of the genome by viral polymerase, translation of viral mRNA, replication of viral genome, assembly of progeny particles and finally budding of the mature virion (Figure 2.8).

2.4.1 Adsorption

The first step in infection is the binding of the rhabdovirus to the target host cell surface receptor. Adsorption is mediated with the help of G protein as the removal of spike protein from the virus reduces its infectivity [66]. Schlegel and co-workers in 1983 showed that the membrane extracts of vero cells could completely inhibit VSV infection. They had also observed that specific compound from the membrane extract responsible for the inhibition was resistant to neuraminidase, trypsin and heating to 100 °C but was soluble in chloroform-methanol and sensitive to phospholipase C [67]. On further experimentation it was concluded that among all the phospholipids, phosphatidylserine (PS) inhibited VSV infection [67]. This indirect evidence suggested that PS could function as a receptor for VSV G protein. Later studies reported particular heptad repeats in VSV G protein that binds to PS [68] and also identified a 19 amino acid peptide from VSV G that binds to PS [69]. However, these studies had not examined the binding of actual virions to normal cell membranes [70].
Moreover, PS is maintained on the inner leaflet of the plasma membrane by aminophospholipid transporter, which specifically transports PS from the outer leaflet to the inner leaflet. PS that appears on the outer leaflet is a signal for engulfment by macrophages or apoptosis and thus has a very short half life. The lack of PS on the cell surface counters the hypothesis that PS serves as a virus receptor. Furthermore, Coil and Miller in 2004 demonstrated that there was no correlation between the cell surface PS levels and VSV infection by using Annexin V which binds to PS tightly and specifically did not inhibit binding of VSV G. Moreover addition of PS to cells did not allow increased virus entry [70]. Thus the overall role of PS in VSV infection is not yet clear.

2.4.2 Entry and uncoating

The virus enters cells through clathrin mediated endocytic pathway [71]. The G protein attaches the virus to host cell receptor, inducing endocytosis of the virion. In the endosome, the acidic pH induces conformational changes in the glycoprotein trimer, which triggers the fusion between virus and cell membrane. Either simultaneously or sequentially, the ribonucleoprotein (RNP) core dissociates from the matrix protein and completes the uncoating process. The dissociation is achieved by exploiting the viroporin activity of the G protein. When the virions are exposed to low pH environment within endosomes during entry, the conformational changes of the G protein induces pore formation on the viral envelope, which allows the protons to enter the virion. Virion acidification triggers conformational changes in the M protein that reduces the interaction between adjacent M protein monomers and the interaction between M protein and the RNP skeletons [48] thus releasing the RNP into the cytosol of the infected cell.
**Figure 2.8:** Rhabdovirus life-cycle

The steps illustrate virus adsorption and endocytosis, envelope fusion with endosomal membranes and subsequent release of nucleocapsids containing parental genomes into the cytoplasm. Released RNA undergoes primary (1°) transcription, genome replication to produce nucleocapsids containing antigenomes and progeny genomes and secondary (2°) transcription which is followed by packaging of progeny genome, virion assembly and budding from host plasma membrane.

### 2.4.3 Transcription

The released RNP serves as a template for transcription of the viral genomic RNA by the associated RdRp (L and P proteins) hence establishing a productive infection. If the process of dissociation between M protein and RNP fails, then a successful infection cannot occur as the M protein is shown to be an inhibitor of viral transcription [48]. *In vitro* studies have proven that the M protein is in dynamic equilibrium with the RNPs and the concentration of M protein determines its reversible association or dissociation with the RNPs. Low concentration of soluble M protein in the cytoplasm of the infected cell favours the process of dissociation while a higher concentration favours the condensation of RNPs [48]. Virus polymerase composed of L protein and phosphorylated form of P protein transcribes the genome RNA with the progressive attenuation at each intergenic region to synthesize the leader RNA and five capped and polyadenylated mRNAs. The viral genes are transcribed in a sequential manner as the polymerase enters a single site at the 3’ end of the genome and also the genes are formed in decreasing amounts with the increasing distance from the 3’ promoter in an order N, P, M, G and L indicating polar transcription [72, 73, 74, 75]. It has been
suggested that RdRp remains associated with the N-RNA during transcription of genomic template and reinitiates the synthesis of downstream genes upon termination [76].

### 2.4.4 Replication

During replication the polymerase read through the termination signals present at the intergenic regions to eventually duplicate the entire genome. N protein is recruited on nascent genome RNA to protect it from cellular RNAse action as it is synthesized [7]. Unphosphorylated form of P protein modulates the polymerase activity and brings about anti termination. During transcription and replication the L protein is positioned between the two lobes of the N protein where the RNA lies to gain access to the genome. The L protein cannot bind to the RNA on its own and this binding is made possible by the mediation of P protein. However, the L protein binding to P protein is not dependent on the phosphorylation of P. The N protein undergoes some conformational changes in order for the L protein to gain access to the genome. The dimeric P protein assists L protein to associate with the viral genome. This dimer associates with both the L protein (enabling L to associate with the viral RNA) and the N protein. N protein is enforced to open momentarily and expose the RNA as the polymerase passes along. A destabilization event arose due to P-L binding has been identified as a cause for triggering the mechanism for forcing the conformational change that culminates in the opening of N. During the replication process, RNA is temporarily unencapsidated to allow the polymerase to use it as template. As the viral polymerase synthesises the RNA, the exposed region of RNA is encapsidated and the next stretch becomes available for synthesis [77].

### 2.4.5 Assembly and Budding

Assembly and budding is a defined process which occurs at specific membrane locations within the cell. Viral components, majorly, glycoproteins contain specific signals which aid in targeting the budding sites. Budding in rhabdoviruses primarily takes place at plasma membrane but it is not the default site as was thought primarily. Cellular proteins and lipids localized at virus budding sites were found to be the constituents of cellular budding machinery and thus altered the view of plasma membrane being the default budding site [78]. Virus assembly occurs at cell surface as a result of specific interactions between G and M proteins in association with the nucleocapsid. According to the model described by Jayakara and co-workers, the first step is the formation of nucleocapsid core through binding of N protein to genomic RNA. While RNP’s are being formed in the cytoplasm, G protein localizes
to sites favourable for initiation of budding. When sufficient amounts of M protein have accumulated in the cytoplasm and a sub-population has concentrated in the inner leaflet of the plasma membrane, nucleocapsids localize to the plasma membrane and are condensed into tightly coiled structures (skeletons) by interaction with M protein. Interaction of M–RNPs with the bud site (containing locally high concentrations of G protein) and the progressive condensation of the RNP core results in evagination of the membrane which envelops the underlying condensed RNP core. Thus, condensation of RNPs by M protein results in formation of bullet-shaped protrusions extending from the plasma membrane. Cellular components such as Nedd-4 associate with the PY motif of M which either directly or indirectly result in release of mature virions [79]. Although this model incorporated all the details of the research done in this direction but still some questions regarding the site of budding, requirement of host factors in rhabdovirus budding and the factors required for initiation of budding remain unanswered till date.

2.5 Diagnosis and treatment of CHPV mediated disease

Keeping in view the continual outbreaks and the short window period to act against the disease, diagnostics play an important role for the detection of the virus. Several tools are now available for the diagnosis of CHPV infection. During the outbreak of 2003, an ELISA assay was developed by Rao and co-workers for the detection of anti-CHPV IgM antibodies in the host sera [4, 5, 24]. However, considering the rapid course of the disease progression, the detection of viral RNA soon emerged as the method of choice. In this direction, hemi-nested RT-PCR [4] and nested RT-PCR [5] were used for the detection. The problems of detecting the low titers of the virus in the blood and the cross contamination observed with nested RT-PCR necessitated the need for the development of Real Time PCR based method [80]. This method had an advancement over conventional RT-PCR based detection systems in being rapid and reproducible. The technique also allowed the quantification of the viral RNA and greatly reduced the contamination problem.

In the absence of any antiviral and vaccines, no specific treatment is available for the disease. Children detected with CHPV infection are treated only symptomatically. Diagnostics help in the detection of the viral infection but in the case of CHPV infection as the window period between the detection of clinical symptoms and fatality is merely 2-3 days, the therapeutics based on vaccines and inhibitor molecules becomes very important. Recently recombinant G protein based vaccine was developed and tested [59, 60].

19
2.6 Interactions among viral proteins

Functional analysis of the viral protein interactions has led to a better understanding of VSV life cycle and pathogenesis. The Nucleocapsid protein encloses the synthesized genomic RNA [29] and protects it from the action of RNases. Its association with the P protein renders it soluble and active by preventing the formation of N protein aggregates. Furthermore, this association also confers specificity for RNA-binding activity [81]. N protein also associates with M protein and this association is important for encapsulating ribonucleoprotein (RNP) cores. D’agostino and co-workers proved that disrupting this interaction can impair the viral assembly [82]. P protein is another vital regulatory protein that plays an essential role in transcription and replication. It also facilitates the binding of L protein with the N-RNA template to form a tripartite complex for genome transcription [9]. The matrix (M) protein of VSV, an important structural component of the virus plays a central role in viral assembly and virion budding by connecting the viral envelope and the RNP core [83]. Interaction of M protein and G protein has been observed in VSV and this interaction is important for clustering of G protein in the regions of viral assembly [84]. Self associations among the viral proteins are also essential for viral life cycle. Dimerisation of M protein facilitates long range organization of M molecules [85] essential for viral budding. Self-association of P protein is crucial for transcriptional activity [86] and G protein also self associates in form of trimers that act as viral spike [87]. In CHPV context, the only known protein interactions are of N and P protein and their self associations [19, 20, 88].

2.7 Importance of protein-protein interactions

Protein-protein interactions (PPIs) are essential for many biological functions and the knowledge of PPIs, help understanding cellular processes [89]. Association or dissociation of protein complexes exhibits the cellular response to perturbations (internal or external) occurring in signal transduction pathways. Transient associations between proteins form the basis for processes like signal transduction, hormone–receptor binding, recognition of antigen by antibodies and enzyme inhibition. Stronger interactions are required for those proteins where the multimeric state helps in determining stability and function such as oligomeric enzymes and structural protein assemblies. The identification of protein interactions can help in providing crucial information about the molecular functions of unknown proteins. Proteins interact with each other in a highly specific manner, and these interactions play a key role in many cellular processes; in particular, the distortion of protein interactions may lead to the
development of many diseases. Recently, protein interaction studies have been carried out for many viruses like Gamma Herpes virus MHV-68 [90], Kaposi’s Sarcoma Associated Herpes virus [91], Epstein Barr virus [92], SARS-Coronavirus [93] and Measles virus [94]. These studies have contributed towards understanding the functions of uncharacterized proteins, their role in viral infection, their mode of action in the course of viral growth, propagation, structure formation and particle assembly.

One of the potential applications of protein interaction analysis is in therapeutics. Due to the specific nature of PPI inhibitors, they target particular interactions without causing any interference with host proteins and enzymes. Evidence is available on how small molecules modulate PPIs directly by blocking the interaction surface or indirectly by binding to an allosteric site to induce a conformational change in the protein [95]. Entry inhibitors that block HIV and host receptor binding are already in clinical use [96, 97]. It is noteworthy that most of these drug molecules are active against strains of HIV resistant to protease and reverse transcriptase inhibitors. Recently, in the direction of developing therapeutics against poxviruses, the compounds that inhibit the protein-protein interactions between D4 and A20 have been identified [98]. These compounds have been shown to bind D4 and exhibit antiviral activity.

2.8 Methods for detection of protein-protein interactions

The initial step of any protein-protein interaction study is the screen for novel binding partners for the protein in question. A range of methods have evolved in order to detect, analyse, and quantify protein interactions. Some of the commonly used methods are discussed here.

2.8.1 Yeast two-hybrid system

The yeast-two hybrid (Y2H) system was developed by Fields and Song in the year 1989 [99]. They generated a new genetic system to study protein-protein interactions employing the yeast strain, *Saccharomyces cerevisiae* and characteristics of its transcriptional factor GAL4. The GAL4 protein consists of two separate functional domains: an N terminal domain that binds specific DNA sequences (binding domain/BD), and a C terminal domain comprising of acidic regions required to activate transcription (activator domain/AD). The GAL4 DNA-binding domain (BD) was combined with a protein “X” (bait) and the GAL4 activating region (AD) to a protein “Y” (prey) to generate a system of two hybrid proteins. It
was based on the principle that protein X and Y can form a complex that will bring the two GAL4 domains close to one another initiating the transcription of the reporter gene governed by GAL4. In order to test the system, two distinct plasmids were constructed carrying sequences coding for GAL4 BD and GAL4 AD fused in frame to genes coding for proteins “X” (bait) and “Y” (prey), respectively. These two plasmids were introduced into a yeast strain containing the reporter gene, LacZ, fused to the GAL1 promoter. The growth of blue-colored yeast colonies on the medium supplemented with a chromogenic substrate [99], 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) indicates β-galactosidase activity resulting through binding of BD and AD fusions complex to GAL1 promoter.

Y2H uses transcriptional activity as a measure for detection of protein-protein interactions. Various interpretations of the basic two-hybrid system have been reported in literature [100, 101]. The frequently used version utilizes the selectable reporter gene GAL1-HIS3. When the binding domains and activating domains fusion proteins interact, the reporter gene, HIS3 is activated resulting in the yeast cell to grow on medium lacking the amino acid histidine (Figure 2.9). Y2H has emerged as a powerful technique due to its ability to identify an interacting protein and the gene which encodes for it. This technique can be used to detect interactions between candidate proteins whose genes are available by constructing the appropriate hybrids and testing for reporter gene activity [102]. Small domains of the interacting proteins integral for the interaction can be identified by deletions in the DNA encoding one of the interacting proteins [100]. Additionally, point mutations can be tested to identify specific amino acid residues critical for the interaction [103]. Most importantly, the two-hybrid system can be used to screen cDNA libraries of activation domain hybrid plasmids to identify protein partners that bind to a protein of interest.
Y2H is sensitive towards detection of weak and transient interactions. Such interactions may not be detected using yeast based assays, but may be critical for functioning of complex biological systems. The experimental setup of Y2H is comparatively simple and it follows a particular protocol that allows for the simultaneous detection and characterization of interactions. The experiment is carried out in the yeast cell, thus enabling the protein of interest to fold in the near native conformation and thereby helping the detection of genuinely interacting proteins. However, Y2H system has specific limitations that include false-positive and false-negative results. False-positives can be generated due to the interaction of the proteins with the promoter sequence or the DNA upstream to the reporter gene. False-negatives can arise due to improper folding of the fused proteins, their inability to localize in the yeast nucleus, inappropriate post-translational modifications and proteins not being functional when expressed as a fusion protein. Hence other methods are used in parallel to validate the two-hybrid results.

2.8.2 Detection of protein interactions using fluorescence based methods

With the introduction of new imaging technologies, development of genetically encoded fluorescent proteins (FPs) and the increasing capability of software for image acquisition and analysis, the study of protein functions and their interactions in native environment have become easier. The most popular methods employed are; filter-based FRET (ratio-imaging/sensitized emission) and lifetime measurements (fluorescence lifetime microscopy, FLIM). Fluorescence resonance energy transfer (FRET) measures the mechanical
effect between a given pair of fluorophores, i.e., a fluorescent donor and an acceptor, where, with the excitation of the donor, energy is transferred from the donor to the acceptor through dipole-dipole coupling (resonance) [104]. FRET is characterized by the efficiency of the energy transfer and measures the fraction of the photons absorbed by the donor and transferred to the acceptor. This technique is widely used due to the high specificity and sensitivity in quantification of various molecular parameters of different biological processes in vivo, mainly molecular interactions among proteins [105]. It is an important tool for studying molecular interactions due to the fact that the range at which energy transfer occurs is approximately 10 nanometres and the transfer proficiency is sensitive to the distance between the fluorophores. This method has been used to investigate interaction between Gag proteins of Rous sarcoma virus in living cells [106]. FLIM, although an extension of FRET has the advantage in being independent of the fluorophore concentration, as this technique does not include the measurement of the number of photons emitted rather it measures the relaxation time of the excited fluorophores [107, 108].

2.8.3 Tandem affinity purification (TAP) method

Tandem affinity purification tag [TAP tag; 109, 110] is a multi tag consisting of a protein of interest, a calmodulin-binding peptide, a TEV protease cleavage site, and protein A for immobilization. This tag is utilised for rapid purification of protein complexes [111] and has application in the detection of protein-protein interactions as well. The open reading frame (ORF) of the target protein is fused with the DNA sequences encoding the TAP tag and is allowed for expression in a host cell. Extracts prepared from these cells are then used to recover the target protein and the interacting partners in two-step process. In the first step, protein A binds to an IgG matrix and after washing out the contaminants, the protease cleaves the link between protein A and IgG matrix. The eluate of the first step is then incubated on calmodulin-coated beads in the presence of calcium. After washing, the target protein complex is eluted. The components of the complex are screened by polyacrylamide gel electrophoresis and the fragments are identified by Mass Spectroscopic analysis. TAP-MS can report on higher-order interactions beyond binary and, therefore, provides direct information on protein complexes [112]. More than 200 distinct protein complexes in yeast have been identified, characterized and validated by TAP [113]. Transient interactions and complexes, however, cannot be detected by TAP technique.
2.8.4 GST pull-down assay

Glutathione S-transferase (GST) pull-down assay is amongst the most commonly used methods for validation of interactions observed through other systems. GST pull-down uses a GST-fusion protein (bait) bound to glutathione (GSH)-coupled beads to affinity purify any proteins (prey; with a fusion tag other than GST) that interact with the bait. Fusion proteins are expressed in protease-deficient strains of *E. coli*, and the protein lysates are prepared for the pull-down assay. The GST tag of the fusion protein enables it to bind to the glutathione immobilized sepharose beads. Upon interaction of two proteins with each other, a complex is formed on the glutathione sepharose beads (Figure 2.10), which can be eluted with free glutathione [114, 115]. The eluted complex and thus the interaction among the two proteins is analysed by western blot using tag specific antibodies. The technique has been utilised for the identification of associations among various viral proteins in different protein interaction studies [116, 117].

![Figure 2.10: Schematic representation of GST pull-down assay](image)
2.8.5 Co-immunoprecipitation

Co-immunoprecipitation (Co-IP) is one of the most commonly used techniques which utilize target specific antibodies to capture the interactors from whole-cell extracts [118]. An antibody against a target forms a complex with it and the complex is later captured and purified using antibody-binding molecule immobilized on a solid support. Unbound proteins in sample are washed away and bound protein complexes are eluted followed by their analysis on SDS-PAGE and western blot. For elution of interacting proteins (secondary target or prey) from a sample, it is imperative for the primary target (or bait) to be a known protein. The antibody used in the reaction is raised against this primary target. Analysis of interacting protein complexes also helps in understanding binding affinities of the protein partners involved as well as the kinetics of their binding. The identification of interactors is generally followed by a series of biochemical and genetic tools for their characterization as well as to understand their role in the biology of the bait protein. However, high background noise from non-specific interactors and antibody interference often creates a problem in identification of true putative interactors. These issues are being addressed by use of advanced techniques like cross-linking. Many of the viral-viral and viral-host protein interaction studies involved the usage of Co-IP [16, 93].

2.8.6 Protein microarray

With the advancements in DNA-based array technology, the transition of protein microarrays has been encouraged by implementing various software tools and proteomics based technology platforms. Protein microarray has recently become fundamental for a variety of applications including high-throughput proteome analysis. A protein microarray utilizes an array substrate that includes silicon, glass, plastic and synthetic polymers and to this surface an inorganic layer such as a metal oxide or a noble metal is applied [119], on the top of this is an organic layer to which a library of recombinant proteins is immobilised directly or through an affinity tag [120]. The proteins are oriented in such a way that their active site faces away from the organic layer for interactions with other proteins in the biological sample [119]. Once the proteins are immobilised, a complex mixture of proteins labelled with fluorescent dye is added to the array. If there is any interaction between the analyte proteins and the immobilised protein, a fluorescent signal is emitted that can be read by a laser scanner. Interactions between the proteins can also be detected by luminescence [119].
Though protein microarrays are very useful, their widespread use is relegated due to the obstacle of producing large amount of pure recombinant proteins [121]. The protein arrays also need the complex and expensive equipment for handling of samples and their detection, though the technology is evolving at a fast rate [122]. Protein microarrays have been employed for analysing the HIV glycoprotein (gp120) interactions with host factors [123].

2.8.7 Phage display

Phage display, first described in 1985 by George P. Smith, is a high-throughput technique to study protein-protein, and protein-DNA interactions by allowing expression of polypeptides on the surface of phage particles. M13 and fd filamentous phages are most commonly used to display foreign proteins or peptides [124].

Expression of exogenous protein on phage surface is achieved simply by creating a fusion between the foreign gene and the viral coat protein gene. Foreign polypeptide libraries with diversities of millions and billions can be screened at once using bio-panning (i.e., affinity purification). One cycle of bio-panning involves the amplification and production of phage particles; exposure of phage particles to the target; followed by washing and subsequently, elution of bound phages. The eluted phages are used to infect the bacterial host, and hence are amplified. A typical screening procedure requires these bio-panning rounds to be repeated three to six times. Phage display libraries of natural and synthetic peptides, proteins, protein domains and synthetic antibodies are now being constructed on routine basis [125, 126, 127]. In a recent study, nucleolar and coiled-body phosphoprotein 1 (NOLC1) was found to interact with NS1 of H5N1 upon screening of human liver cell cDNA library displayed on T7 phage [128]. Phage display has now become a popular technique for elucidation of protein interactions. Another major application of phage display is directed evolution of proteins which involves manipulating a population of peptides towards desired properties by creating random sequence variation. Phage display of antibodies has evolved to be an effective tool in generation of antibodies as it does not require the target antigens to be immunogenic. Phage display has limitation that it requires the purification of target protein. Overall, combining phage display with other high-throughput procedures can give rise to new applications in future.
2.8.8 *In-silico* prediction and validation of protein-protein interactions

As genome sequences are being continuously revealed and the amount of interaction data is increasing, a growing number of computational methods for predicting protein interactions are emerging. These methods compare the presence or absence of genes within genomes, the spatial relationships among genes, gene fusion events, and co-evolution among protein pairs to assess the possibility of protein interactions. Due to the limitations of experimental methods of binary interaction and complex determination, there is a need of computational models to mend the continuously increasing gaps among protein annotation and genome sequence information. The *in silico* methods for protein-protein interaction prediction such as gene fusion [129], gene neighbourhood [130] and phylogenetic profiles [131] predict functional associations between proteins. The gene fusion approach, for instance, suggests that if one gene product in one organism appears to be expressed as two separate gene products in a second organism, a functional association between the two gene products is implied in that second organism. Although this can suggest a physical interaction between the proteins concerned, links between molecules may be less direct. Recent advances in the prediction of new physical associations between proteins have adopted a threading based approach to predict protein–protein interactions [132, 133]. Furthermore, the prediction of protein interactions can be analyzed if the structure of two proteins is known. To confirm data resulting from large-scale studies, bioinformatics methods are often used [113, 134]. By the application of different algorithms, results are integrated with available data on expression, localization, function, evolutionary conservation, protein structure and binary interactions.