Introduction
1. General Introduction

1.1 RNA viruses

The pathways of genetic information flow in cells, codified as the “central dogma” of early molecular biology, are the replication of DNA from DNA, transcription of RNA from DNA and the translation of proteins from RNA. Another pathway outside this scheme that is attracting growing attention is the copying of RNA from RNA, mediated by enzymes called RNA dependent RNA polymerases. It’s believed that these enzymes must have played a very crucial role in the early stages of evolution when RNA was the primary genetic material. These enzymes also have crucial roles in contemporary biology and these enzymes are encoded by a wide variety of RNA viruses and they are absolutely imperative for the life cycle of these viruses. RNA viruses are the most abundant molecular species infecting humans, animals and plants. A comparative analysis of the structures, genetic organization and replication pathways indicate that RNA viruses use disparate strategies to ensure their multiplication in cells and their stability as free particles. About 50 new viruses have been recognized as emergent in the past two decades. Most of them are RNA viruses. A few examples of medically important RNA viruses are Ebola virus, which causes lethal hemorrhagic fevers; influenza virus, which in a single 1918-1919 pandemic killed tens of millions of people, and hepatitis C virus, which is carried by over 150 million people can cause severe liver damage and cancer. RNA viruses are classified into three major classes categorized by whether the infectious virion particles contain the genome as dsRNA, positive stranded (messenger-sense) RNA, or negative-stranded RNA. The negative sense RNA virus are further classified as Arenaviridae, Bunyaviridae, Orthomyxoviridae, Mononegaviridae, Tenuviridae. The Mononegavirales include, Rhabdovirus, Paramyxovirus and Filovirus. A broad array of human and animal pathogens is included in each of these families (Wagner 1987, Kingsbury 1991).

Rhabdoviruses are widely distributed in nature, and there are about 100 rhabdoviruses – a family of viruses found in plants, reptiles, fish, crustaceans, and mammals. (Banerjee, 1987a,b) Rhabdoviruses generally cause acute infections and can persist under certain conditions for long periods (Wagner et. al., 1987). Their common characteristic is the rod shaped morphology (Greek rhabdo = rod shaped), which is revealed by electron microscopic studies. Rhabdoviruses have linear single strand non-segmented genome having negative polarity. The genome RNA is non-infectious and complementary
to functional virus specific, positive sense messenger RNA. One of the unique features of the negative-strand genome RNA of a rhabdovirus is that it serves as the template for both transcription (mRNA synthesis) and replication (genome RNA synthesis). Moreover, due to the packaging of a novel, virus specific RNA polymerase within the mature virion particle, these viruses serve as one of the model systems to study the transcription, replication and their regulation in vitro. Two genera of rhabdoviruses affect mammals are the Lyssavirus, which invade the central nervous system and the Vesiculoviruses, which invade epithelial cells, usually the tongues to produce vesicles. The major virus in the Lyssavirus family is the Rabies virus, which causes encephalitis in a variety of domestic and wild animals and also to humans. The major virus in the Vesiculovirus genus is the Vesicular Stomatitis Virus (VSV), causes an epidemic but self limiting vesicular disease of cattle in the western hemisphere. The plant and fish rhabdoviruses undoubtedly comprise at least one more genus. There are few unclassified rhabdoviruses also which are not related serologically to other viruses and are antigenically ungrouped. The fish rhabdoviruses are divided in two subclasses according to their host specific origin, namely salmonoid and non-salmonoid. Other vesiculoviruses include Piry virus, Chandipura virus and Isfahan virus.

1.2 Chandipura virus

In 1964 and 1965 an epidemic due to Dengue virus and Chikungunya virus infection occurred in many cities in India (Rao, 1966). One unknown agent was isolated from the blood samples of two infected human female patients in a village called Chandipura near the city of Nagpur in western India and this agent showed cytopathic effect on the BS-C-1 cells in tissue culture. The cytopathic effect was found to be very different from that of other arboviruses isolated and studied in India (Bhatt and Rodrigues, 1967). The suggestive evidences provided by Bhatt and Rodrigues demonstrated that the isolated agent was an arbovirus new to India. The virus was later named as Chandipura virus or CHP virus after the name of the village. From the structure and organization of the genome the virus was found to be a rhabdovirus and was related but not serologically identical to VSV, which is a member of the vesiculovirus groups of the rhabdoviridae family (Dragunova and Zavada, 1979).
1.3 Biological characteristics

1.3.1 Transmission

Like all rhabdoviruses, CHP virus has a wide host range including vertebrates and insects. In nature, this virus has been isolated from mammals, blood sucking arthropods, e.g., mosquitoes, sand flies etc. (Rao et al., 1967) determined the vector potential of mosquitoes for the virus. CHP virus was found to be transmitted by four species of mosquitoes they tested. Among these four species, *Aedes aegypti* and *Aedes albipictus* were found to be more efficient in transmitting the virus than the *Anopheles sthifensi* and *Culex tritaeniorhynchus*. *Culex flagons* pick up the virus by feeding, retain it and are thereby unable to transmit it. Later on, in 1991, Ilkal et al., have shown that *Culex tritaeniorhynchus*, *Cx. bitaniorhynchus*, *Cx. quinquefasciatus* and *Aedes aegypti*—all have vector potential for transmitting the CHP virus to the laboratory strain mice and among them the *Aedes aegypti* was more susceptible. (Tesh and Modi, 1984) demonstrated that, CHP virus could grow even up to four logs within 24 hours intrathoracic inoculation in sand flies (*Phelebotomine* *papatasi*) and by transovial transmission; the virus can be transmitted into next progeny. These phenomena suggest that insects can act both as host and vector. Mosquitoes and drosophila have been found to develop paralytic and/or lethal sensitivity when infected with CHP virus in the laboratory (Brown and Crick, 1979). Available evidences indicate that the mammals may act as dead end host because horizontal transmission is rare and viremia is transient (Barik and Banerjee, 1994).

1.3.2 Clinical symptoms

Clinical data were obtained from the patients whose sera were collected for the isolation of CHP virus. The symptoms were compatible with Dengue and Chikungunya virus infection. The patients showed sudden onset of fever without chills or rigor (100° to 105°F approximately) with pain in the small joints of feet. Later pain in the total body, headache, vomiting also developed; weakness still persisted even after eleven days of illness (Bhatt and Rodrigues, 1967). In a specific case, the person with CHP virus infection was found to be associated with acute encephalopathy syndrome (Rodrigues et al., 1983).

1.3.3 Epidemiological study

A seroepidemiological study was carried out with the blood samples collected from different places of India in the span of 1955-1965 by Bhatt and Rodrigues (1967). Apart
from the negative results from North India and Kashmir, the percentage of sera, which neutralized the virus, ranged from 10 to 89.

1.3.4 Life cycle

For infection of *rhabdoviruses* in the vertebrate cell, the first event is adsorption. Adsorption seems to be mediated through G protein, projected from the virus membrane. It was found that G-protein fuses only with the acidic phospholipid (phosphatidylserine or phosphatidic acid) at acidic pH (Edileman *et al.*, 1984). For infection to proceed the virus should penetrate into the cell. The intact virus enters the cell by means of clathrin-coated pits progressing to coated vesicle and then to larger collecting vacuoles and finally secondary lysosomes. The first virus directed metabolic event after penetration observed is the uncoating of the nucleocapsids and their release in the cytoplasm. The released nucleocapsid starts primary transcription with the help of virion associated RNA dependent RNA polymerase. The positive stranded leader RNA directly goes to the nucleus and shuts off the initiation of host rRNA synthesis (*Wagner et al.*, 1987, 1991). The viral mRNAs are translated into corresponding viral protein by host ribosomes (*Hsu et al.*, 1979). After accumulation of a certain amount of viral proteins the viral polymerase is shifted to the replication mode to synthesize full length positive sense genomic RNA which in turn acts as the template to synthesize further negative sense genome RNA. For assembly and maturation of the progeny virion, protein synthesis from amplified RNP is required. The process by which mRNAs are synthesized from this amplified RNP, is called secondary transcription. G protein is synthesized from mRNA on endoplasmic reticulum membrane associated polyribosomes. It enters the endoplasmic reticulum by means of a signal sequence. After stepwise glycosylation it migrates to the cytoplasmic membrane for fusion and insertion (*Rose and Schubert*, 1987). The other four mRNAs are translated on the cytoplasmic polyribosomes from monocystronic messengers. The M protein binds to the progeny nucleocapsids that resulted in the formation of a tightly coiled skeleton of final nucleocapsid structure of the *rhabdovirus* virion (*Newcomb et al.*, 1982). The nucleocapsid-M protein complex migrates to the region of the cytoplasmic membrane that contains the newly inserted but randomly distributed viral glycoprotein. The nucleocapsid-M protein complex appears to bind at the surface of the membrane rich in G protein and phosphatidylserine (*Pal et al.*, 1987), enveloped and buds out as mature virion.
13.5 Effect of virus infection on host protein synthesis

After infection of VSV and other rhabdoviruses, the viral replication takes place in cell cytoplasm, resulting in rapid inhibition of host cell gene expression (Wagner et. al., 1987, 1991; Thomas, 1985). The earliest observed effect is the cessation of maturation of small nuclear RNA (snRNA and URNAs). It is proposed that this block in snRNA metabolism results from the inhibition of export of snRNA precursors out of the nucleus. The putative viral inhibitor of RNA export would likely be contained in the infecting virions, since the effect of snRNA metabolism are extremely rapid and do not require formation of functional viral mRNA. The prime candidate for such an inhibitor is the matrix (M) protein of VSV, present in 1600 to 1800 copies per virion; synthesis of this protein in the transiently transfected cells interferes with the nuclear gene expression. It has been suggested that the short 47nt leader RNA that is made early in VSV infection plays a role in host cell protein synthesis shut off, but evidence that it directly affects nuclear gene expression is lacking.

1.4 Virus structure and Genome organization

Like all other rhabdoviruses, CHP virus is bullet shaped negative stranded virus surrounded by lipoprotein envelope, which encloses a helical nucleocapsid with a non-segmented single strand of RNA that cannot serve as messenger RNA (mRNA). Electron microscopy studies show that CHP virus is of about 180nm in length and 65nm in width.

The virus consists of two structural components:
(i) The outer membrane derived from host cells.
(ii) The internal ribonucleoprotein particle.

1.4.1 Membrane

The membrane contains 50% lipid and two proteins. The lipid components are derived entirely from the host cells but are selected in somewhat different proportions than those in host cell plasma membrane; the main lipid differences in VSV membrane are the larger proportion of cholesterol and, among the phospholipids, much less phosphatidylcholine compared to sphingomyelin and larger amounts of amino phospholipids. (Pal et. al., 1987) The protein components are the integral glycoprotein (G), which protrudes externally and the peripheral matrix (M) protein, which lies in the inner bilayer of the membrane. The G protein and the M protein both have marked affinity for association with the lipid vesicles (Pal et. al., 1987). The carboxyterminal hydrophobic domain
of the G protein transverse into the neutral lipid vesicles whereas the peripheral 
and positively charged M protein attaches superficially only to the membranes con­
taining the negatively charged phospholipids. Both G and M protein characteristi­
cally alter the dynamic properties of the interacting lipid bilayers, but in quite 
opposite ways. The effects of these proteins on the membrane can be interpreted by 
the finding that VSV (IND) virion contains 1205 molecules of G protein and 1,820 
molecules of M protein per virion (Thomas et. al, 1985).

1.4.2 The ribonucleoparticle

The nucleocapsid or the ribonucleoparticle (RNP) is believed to be the infec­
tious component for all rhabdoviruses. The ribonucleoparticle consists of the 11.6kb 
single stranded genome RNA, enwrapped by about 1200 molecules of the Nucleo­
capsid protein N. Two other minor proteins called the Large protein L and the 
Phosphoprotein P are also tightly associated with the ribonucleoparticle. This RNP 
complex along with L and P proteins consists of the infectious core of the virus. 
Inside the virion the RNP exists as a helically coiled structure which is held in place 
by the matrix protein M by it's interaction with the lipid envelope.

1.4.3 Genome RNA

The organization of the genome RNA of almost all Rhabdoviruses is extremely 
tight and compact. The genomic RNA is of negative sense, and is noninfectious. It 
needs RNA dependent RNA polymerase to synthesize the positive sense RNAs to 
become infectious. The idea that the genome of the rhabdoviruses are simple and 
efficient for transcription came from the information that 99.37% of the nucle­
otides are transcribed and 93.88% are translated into proteins (Abraham and 
The average length of genomic RNAs of all rhabdoviruses so far sequenced is around 
11kb. Five genes for five viral proteins, N, P, M, G, and L are present sequentially as 
written, on the genome from the 3' to the 5' end. N gene is preceded by a small gene 
for leader RNA of around 50 nucleotides in length. Leader RNA as long as 140 
bases was also isolated in Sonchus Yellow Net virus. The leader RNA does not 
code for any protein but is very important for viral replication as it interacts with 
the N protein for faithful end-to-end copying of the genome during replication and 
also proper encapsidation of the replicated genome. Thus evolutionarily the N gene
could not be located at the exact 3' end of the genome, since in that case the transcribed N-mRNA had to function both as a message as well as it has to interact with its own encoded polypeptide, namely the N protein. So to circumvent this problem the short leader sequence is positioned preceding the N gene at the extreme 3' end of the genome, which can as well serve as to introduce a strong promoter element for RNA polymerase to initiate transcription and replication. The sequences in the junctions between the genes are more or less conserved in case of the vesiculoviruses, except for the starting sequence of the M gene of CHP virus. In case of rabies virus the intergenic regions are variable both in length and the nucleotide sequence (Tordo et al., 1986, 1988). The presence of common sequences (5'CUGUUAGU,CAUA3') (Barr et al., 1997) at the intergenic regions gave the idea of common strategies for termination of RNA synthesis, polyadenylation of mRNAs and attenuation of RNA synthesis. The U7 stretch probably is the signal for polyadenylation, where the RNA polymerase incorporates several A residues (Banerjee et al., 1974; Rose, 1981; Emerson, 1982). There are five AU rich sequences present at the strategic regions on the genomes of VSV (Iverson and Rose, 1981, Banerjee et al., 1984; Gill and Banerjee, 1985). This are presumed to be equivalent to the TATA boxes present at the upstream regions of eukaryotic genes and thus seems to be the binding sites for RNA dependent RNA polymerase (Keene et al., 1980; Banerjee, 1987a,b). The untranslated 50nt region at the 5' end of the rhabdovirus genome is thought to contain the necessary signals for replication. This sequence is not fully complementary to the 5' end of the genome and contains residues for the binding of N protein and probably that directs the transcription of uninterrupted full-length genome.

1.4.4 The Nucleocapsid protein (N)

The N protein performs several important functions in the viral life cycle. The N protein encapsidates genome RNA in a precise RNase resistant helical structure that can be compared with histone mediated enwrapping of DNA molecule into nucleosome structure. Only this encapsidated form of the genome can be recognized by viral polymerase as its template during both transcription and replication (Banerjee, 1987). The N protein accounts for 90.7% of the protein mass of the purified nucleocapsid (Thomas et al., 1985). Nucleocapsid proteins not only protect viral genome from RNase action but it is also thought to play some vital regulatory roles, it interacts with the viral
polymerase during genome expression and with the M protein during maturation and budding. The intracellular concentration of unassembled N, which exists as a soluble N°-P complex is believed to regulate the switch from transcription to replication (Blumberg et. al., 1981, Banerjee and Barik., 1992, Lamb and Kolakofsky,1996, Patton et. al., 1984).

Direct hybridization experiments showed little sequence homology between the N-mRNAs of VSV (IND) and VSV (NJ) whereas the 5' hexanucleotide sequence was identical. Both the VSV (IND)-N and VSV (NJ)-N proteins contain 422 amino acid. The longest region of perfect identity between the two amino acid sequences (i.e., VSV (IND)-N and VSV (NJ)-N) is 23 nucleotide long (803-840; 902-915; 971-984). At the level of amino acids, 290 residues out of 422 were found identical which correspond to 68.7% homology. When conservative amino acid replacements are considered, it seems to approach 80% or more. The longest continuous block of identical amino acid residues is 21 (214-234). There are also two blocks of 18 residues and one of 17 residues that are perfectly conserved. Also of the 65 amino acid residues from the position of 264-328 in the N protein sequence, only 4 are not identical. In addition to this, there are several regions having at least 50% homology (Banerjee et. al., 1984). Comparison of amino acid sequences of N protein demonstrates that the number of residues common to CHP and VSV (NJ) but not present in VSV (IND) and that common to CHP and VSV (IND) but absent in VSV (NJ) are about the same, suggesting CHP to be evolutionarily equidistant from VSV (NJ) and VSV (IND) (Masters and Banerjee,1987). Alignment of the known N protein sequences in the Paramyxoviridae family suggests that the N protein is composed of two domains. The N terminal 80% which is thought to constitute a globular domain in the protein structure is well conserved with in a genus, whereas the C terminal 20% although generally negatively charged, is much more divergent (Curran, 1993). This hypervariable C-terminal tail appears to be a tail that extends from the globular body of the assembled N protein. In Sendai virus this C-terminal region is sensitive to trypsin, but in spite of this tryptic cleavage the structure of the nucleocapsid remains unaltered. Since the tryptic cleavage does not alter the overall structure of the nucleocapsids and also it fails to alter the RNase sensitivity of the nucleocapsids, it is believed that the major determinant of the nucleocapsid morphology must reside in the conserved N terminal domain of the protein. Structure function studies in Sendai virus N protein has demonstrated that the C termi-
nal tail is dispensable for encapsidation, whereas the deletions within the N terminal domain abolishes this activity (Curran, 1993, Buchholz and Homann, 1994, Buchholz and Homann, 1993). Although the C-terminal tail of the N protein is not required for the assembly of the nucleocapsids but this C-terminal tail is needed for template functions, because in Sendai virus it has been seen that nucleocapsids that assemble with tailless N protein are unable to serve as templates for further rounds of replication (Curran, 1993). It is seen that the hypervariable C-terminal tail helps to mediate Phosphoprotein binding and hence polymerase binding (Buchholz and Homann, 1994). This template function apparently does not reside exclusively in the C-terminal tail because N proteins carrying alanine substitutions mutations between amino acids 114 to 129 were assembly competent, but these were unable to serve as templates for replication. (Myers and Moyer, 1997). In both VSV (IND) and VSV (NJ) strains it's seen that the C-terminal end is very important for encapsidation. Even deletion of 5 amino acids from the C-terminal end abrogated the encapsidation ability of the N protein. Further mutational studies have shown that the penultimate Lysine residue in the C-terminal end of the N protein plays a very crucial regulatory role in the encapsidation process (Das and Banerjee, 1999).

When expressed in eukaryotic cells the VSV (NJ)-N protein rapidly aggregates in high molecular weight complexes in the absence of VSV genome RNA (Sparague et al., 1983). Electron microscopic studies of VSV infected cells demonstrated that nucleocapsid structure accumulated even in the absence of genomic RNAs. N proteins of Measles virus(Spehner et al, 1991) and Sendai virus (Buchholz et al., 1993) when expressed in mammalian cells, N proteins of Measles virus (Fooks et al., 1993) and RSV (Meric et al., 1994) when expressed in insect cells—all formed nucleocapsid like structures. Even when expressed in E.coli, the Measles virus N protein formed nucleocapsids like structures. (Warnes et al., 1995). All this proves that N protein can self assemble into nucleocapsid like structures, in absence of any other modifications of cofactors that may be present in a eukaryotic cell. The N protein has a strong RNA binding ability, and in absence of any other viral factors it can bind to nonspecific RNAs. But somewhat surprisingly for a protein with such strong RNA binding ability there no classical RNA binding motifs present in N protein. A peptide corresponding to 298-352 amino acids of the rabies virus N protein, which is also the most conserved region in the Rhabdovirus N proteins, showed a very strong RNA binding ability. However, no significant sequence similarity was de-
tected between this peptide and known RNA binding proteins in the database. This indicates that N protein may possess a new type of RNA binding motif. (Kouznetzoff et. al, 1998).

In Rabies virus the N protein undergoes phosphorylation and it was seen that unphosphorylated Rabies virus N protein binds the leader RNA more strongly than the phosphorylated N protein. However the rates of transcription and replication of the rabies virus minigenome were significantly lower with the unphosphorylated N than with the phosphorylated N. The tight binding of RNA by the unphosphorylated N may make it more difficult for other viral proteins to unwind the encapsidated RNA for transcription and replication by viral RNA polymerase L. Thus it appears that phosphorylation of Rabies virus N protein supports viral RNA transcription and replication, possibly by modulating the encapsidation of genomic RNA (Yang et. al, 1999).

1.4.5 The Phosphoprotein (P)

The P protein is so named because of its highly phosphorylated nature. It is increasingly becoming clear that P protein of negative stranded RNA viruses, with a linear single stranded RNA as the genetic material, is an important auxiliary protein with direct roles in transcription and replication of the virus. (Banerjee et. al., 1991, Horikami et. al., 1992). Various important functions of P protein are as follows,—P acts as a transcription factor, when it complexes with the viral polymerase, L, P undergoes differential levels of phosphorylation which is very important for the viral life cycle., P also forms various complexes with N protein and keeps N protein in a replication competent form. The size of P protein varies considerably within the Paramyxoviridae, with the respiroviruses and the morbilliviruses being (507-603) amino acids long, and the rubulaviruses being 245-397 amino acids long and those of the pneumoviruses are the smallest, it being 241 amino acids long. In the Paramyxoviridae sequences of many P proteins are now known. (Barrett et. al, 1985, Berg et. al, 1992, Galinski et. al., 1986, Shioda et. al, 1986, Paterson et. al, 1990). Sequencing has revealed that P genes are polycistronic (except of those of pneumoviruses and hPIV1), expressing proteins from up to three reading frames using different start codons. The respective proteins for Sendai virus are C, C', Y1, Y2, V, W, I and X. There is little information available on the functions of the respective proteins, although evidence suggests that C, V, and W proteins

The phosphoprotein \((P)\) of different \textit{rhabdoviruses} is quite well studied in the recent past. In \textit{rhabdovirus} research considerable amount of new information regarding the phosphoprotein \(P\) function has accumulated in the last few years. (Masters and Banerjee 1987) first sequenced the \(P\) gene of CHP virus. The CHP-\(P\) protein contains 293 amino acid residues in comparison to the 274 amino acid residues of VSV (NJ)-\(P\) (Gill and Banerjee, 1985; Hudson et. al., 1986, Masters and Banerjee, 1987). The calculated molecular weight of CHP-\(P\) protein is 32.5 KDa. Overall the protein is highly acidic like the phosphoproteins of different \textit{rhabdoviruses}, viz., VSV. The N-terminal part is highly acidic in nature and is rich in Asp and Glu residues. Due to the highly acidic nature, CHP-\(P\) protein migrates as 45 KDa proteins in SDS-PAGE, which indicates that SDS cannot neutralize the charge of the \(P\) protein completely. Alternatively, a typical secondary structure of the protein may cause the altered mobility.

Perhaps the most important common feature of the \textit{Mononegaviridae}, (MNV) is that they are all small, relatively simple viruses, possessing no more than 10-12 genes. This extreme parsimony necessitates two general strategies during infection; first that maximal use will be made of host cell machinery and second, that some gene products will need to perform more than one function. Phosphorylation of a viral protein by host cell kinases to produce two or more different forms can implement both of these strategies. As is mentioned previously the most highly phosphorylated protein in any MNV virion or infected cell is generally the \(P\) protein. There are two levels of phosphorylations of the \(P\) protein—first there is a primary phosphorylation by the host cell kinase and then there is a second level of phosphorylation by virion associated kinase, and both these phosphorylation events are very important for the virus life cycle. Mainly serine residues in the N terminal end, of VSV \(P\) and measles virus \(P\) protein (Das et. al., 1995, Hsu et. al, 1982), the middle portion of Sendai \(P\) and hPIV3 \(P\) proteins (Byrappa et. al., 1996, Huntley et. al, 1995) or the extreme C terminal end of the RSV \(P\) protein (Barik, et. al, 1995) have been identified as the main sites of phosphorylation. Some of the cellular kinases involved in the phosphorylation of the MNV \(P\) proteins are —CKII, phosphorylating, VSV, \(P\) protein, RSV \(P\), protein, PKC-\(\zeta\) phosphorylates Sendai \(P\) (Huntley et. al., 1997) and hPIV3 \(P\) protein (De et. al, 1995). From the study with the bacterially expressed CHP-\(P\) it was found that the protein undergoes phosphorylation by a host
kinase that was later identified as Casein Kinase II (CKII) protein (Chattopadhyay et al., 1994). The CKII phosphorylated protein was found to support transcription in vitro (Chattopadhyay et al., 1997). The CKII phosphorylation site was mapped and the site of phosphorylation was detected as Ser62 by site directed mutagenesis (Chattopadhyay et al., 1997). The functional domains of CHP-P protein have not yet been mapped. Interestingly, despite their poor sequence homology, P proteins of various rhabdoviruses have very similar distribution of hydropathicity, suggesting that the overall conformation of the protein may be conserved and is important for its function (Gill and Banerjee, 1985). The N-terminal domain is least conserved and contains the CKII phosphorylation site(s).

In VSV, the P protein was divided into three functional domains. Domain I, which is least conserved and highly acidic in nature comprises of residues 1-137; the domain II comprises of residues 214-247. There is a very poorly conserved region within domain I and domain II (138-213) and is called the hinge region (Gill and Banerjee, 1985). In contrast, domain III (250-274), which consists conserved 25 amino acid residues (Gill et al, 1986). The significance of the spacer or hinge region is thought to provide correct spatial arrangements of the active domains. This is supported by the observations that VSV transcription reconstituted with the two polypeptides representing domain I and domain II and III added in trans is functionally inefficient (Chattopadhyay and Banerjee, 1987, 1988). Such spacer like region is also found in different transcriptional activators of DNA dependent RNA polymerases, viz., Yeast Gal4, GCN4, and HAPI proteins. In case of Paramyxovirus P proteins it was found that the proteins exists as homotrimers, and the oligomerization is mediated by a coiled-coiled motif located within the C terminal end of the molecule (Curran, 1995) In case of VSV P and also CHP the phosphorylation leads to the oligomerization of the P proteins, but in case of Sendai unlike the above two cases P protein oligomerization is not facilitated by phosphorylation. The inherent stability of the oligomeric P protein in Sendai virus is demonstrated by the fact that if the coiled-coiled domain of P-P interaction is deleted then both L-P and N-P interaction is impaired. For VSV on the other hand the monomer–oligomer equilibrium is mediated by phosphorylation of the P protein. In CHP also the oligomerization is mediated by phosphorylation and phosphorylation leads to dimerization of the P protein (Raha et al, 2000). Like Sendai virus, and VSV P protein also exists as trimer in it’s oligomeric state (Curran, 1995b, Gao et al, 1996) In most cases the coiled-coiled region of the P protein is involved in the P-P interaction and it has been found that in Sendai virus (Curran
1995b), RPV P (Shaji et al., 1999). P proteins of negative sense RNA viruses play various important roles. Dimethylsulphate footprinting of VSV transcriptase complex has indicated that a complex formed between L and P is capable of recognizing a 3'-end genomic RNA which can act as a putative promoter (Keene et al., 1981) and such a footprint can be generated by P protein alone. Chemical cross linking studies have shown that the Sendai P protein remains in close proximity with the template RNA (Raghow and Kingsbury 1979), in CHP also it has been shown that P protein interacts with the 3' end leader RNA (Basak et al., 2003). So it is possible that P protein plays a very vital regulatory role in recognizing the start site for transcription and replication. Interaction of P protein with L protein appears to prevent proteolytic degradation of L protein in measles virus (Horikami et al., 1994), Sendai virus (Smallwood et al., 1994) and VSV (Canter and Perrault, 1996). Many of P proteins important functions are mediated by its interaction with N protein. In Sendai virus it has been seen that N-P interaction prevents the illegitimate aggregation of N protein (Curran et al., 1995a). Similarly in VSV the interaction of P protein with nascent N protein prevents the nonspecific RNA binding tendency of N protein (Masters and Banerjee 1988).

Progressive deletion from the C-terminal end of VSV P showed that N-RNA binding function measured either by ultra centrifugation (Gill et al., 1986) or by immunoprecipitation (Emerson and Schubert, 1987a,b), was lost when domain III was removed. However binding as well as transcription was restored in presence of L protein. Furthermore, deletions which remove domain I and domain II also resulted in a decrease in the N-RNA binding and this cannot be restored in presence of L protein. Thus, the highly basic domain III may be an RNA binding domain while domain II may bind L protein to support transcription. An oligopeptide representing the C-terminal 21 amino acid residues was found to inhibit VSV transcription in vitro. It was also found that two serine residues (236, 242) in VSV (NJ)-P in domain II is the site for L associated kinase and mutation of which decreases the N-RNA binding and loss of transcription property in vitro (Chattopadhyay and Banerjee, 1987). By in vitro binding studies it was demonstrated that VSV (NJ)-P protein has at least two sites for L binding under different conditions: (a) in the presence of N-RNA template, domain II binds with L protein to form active transcriptase complex (b) in the absence of N-RNA template domain I binds with L protein (Emerson and Schubert, 1987b).
1.4.6 Large protein (L)

The L protein is the largest protein detected in the negative sense RNA viruses, and it is the least abundant of all the structural proteins (present approximately 50 copies per virion). Within the virion, the L protein remains associated with the N-RNA template through its interaction with the P protein to form the transcribing ribonucleoprotein complex (RNP complex), (Mellon and Emerson, 1978). In case of VSV, genetic studies as well as biochemical analysis using highly purified L and P proteins and the N-RNA template in the in vitro reconstitution assay, demonstrated that the polymerase activity resides in the L protein whereas P acts as a necessary factor in the mRNA synthesis. Consistent with the large size, genetic and biochemical studies suggested that the L protein might also encode the capping, methylation and polyadenylation functions. In vitro reconstitution studies demonstrated that the L protein, although unable to synthesize the full length RNA from the N-RNA template, could initiate RNA chains (De and Banerjee, 1983). In the presence of ATP and [αP32]-CTP, initiated transcripts such as pppA and pppAACA were detected in vitro (De and Banerjee, 1984). The P protein, on the other hand, fails to initiate RNA synthesis, but appears to act at some step subsequent to initiation (De and Banerjee, 1985). The initiation function of L protein and the chain elongation function of P also correlate with other requirements in catalytic and stoichiometric amounts respectively. A protein kinase activity has been found to be strongly associated with the L protein (Sanchez et. al., 1985, Barik and Banerjee, 1991, 1992, Hammond et. al., 1992). Recently the protein kinase activity has been found to be dissociable from the L protein, suggesting the host origin of this kinase (Gao and Lenard, 1995b). Sequence analysis of the L gene from paramyxovirus and rhabdovirus revealed six conserved domains, I-VI (Poch et. al., 1990, Sidhu et. al., 1993a).

Domain I lie in the amino terminal end of the protein (225-416 aas) and have no homology with the currently known motifs. Domain II (503-607aas) just distal to domain I, may be involved in template recognition. Domain III (653-876 aas) contains the putative polymerase active site. The conserved peptide sequence GDN in this region is similar to the conserved catalytic domain GDD of positive stranded RNA virus RNA polymerase (Jablonski. et. al., 1991). Domains IV and V (927-1128 and 1129-1378 aas respectively) have not shown homology with any known motifs; however a putative nucleotide binding motif lies within domain VI (aas 1770-1847). Domain II and domain III of N have been proposed to be the polymarese module
Domain II contains a region, designated premotif A, with two invariant and two highly conserved charged residues when comparisons are made with 23 RNA dependent RNA or DNA polymerases including enzymes from Paramyxoviruses, Rhabdoviruses, Bunya viruses, Arenaviruses, Influenza viruses, Filoviruses, and HIV (Muller et al., 1994). By extrapolation to the crystal structure of HIV reverse transcriptase, two of these residues, R562 and E569 are proposed to contribute to the palm region, and two K453 and R552, are proposed to contribute to the finger region of the Sendai L protein to position the template in the catalytic site. The putative polymerase module in domain II containing the GDN sequence is proposed to lie within the palm region of the RNA polymerase constituting the active site (Muller et al., 1994). Analysis of site directed mutant of these regions of Rabiesvirus L protein (Schnell and Conzelmann, 1995) has shown that GDN is essential for in vivo and in vitro RNA synthesis.

1.5 Viral RNA synthesis

1.5.1 Viral transcription

The naked RNA of Mononegavirales is noninfectious for two reasons: I) the active template is a ribonucleoprotein complex (N:RNA) and II) the genomic RNA cannot serve as a messenger, but must be transcribed by a virally encoded polymerase (P-L complex). Holonucleocapsids carrying L and P proteins can initiate an infection on entry into the cell cytoplasm (Lamb et al., 1976). The polymerase is thought to enter the genome obligatorily from the 3' end of the genome, synthesize the short positive sense leader, terminate, and then reinitiate at the beginning of the first gene. Reinitiation at this first junction is critical because it locks the polymerase, into a transcriptive mode, rather than replicative mode. Polymerases, that read through this junction without concurrently assembling the nascent chain, are non processive and never reach the end of the first gene. Reading through this first junction is generally a rare event. In the pol R mutant of VSV and strain Z of Se V, the percentage of read through is about 20% (Perrault et al., 1983, Vidal et al., 1989). In both these example the phenotype maps not to the viral polymerase but to the N protein of the nucleocapsid, highlighting the fact that this protein plays a crucial and regulatory role in the viral lifecycle. This early event in the lifecycle of the virus is termed as primary transcription and it results in the accumulation of 5' cap and 3' polyadenylated viral mRNAs derived from the infective templates. Pri-
mary transcription starts at the exact 3' end of the genome and does not require any de novo protein synthesis. Kinetic studies demonstrated that synthesis of transcripts follows the order of the genes in the genomes. (Ball and White, 1976, Banerjee, 1987, Iverson and Rose, 1981). In addition there is a gradient in the quantity of transcript following the same 3' 5' order. These observations have led to the so-called stop-start model for transcription, according to which the polymerase enters the N-RNA template exclusively at the 3' end and transcribes the genes successively in the 5' direction (Emerson, 1982). The cis acting transcription signals defining the gene borders have been identified for many non-segmented negative stranded RNA viruses by comparing the sequence present in the genome template and in mRNA ends. The mRNAs start with common short consensus sequences that are templated by the genome, and these are regarded as gene start signals (GS). The first A residue of the poly (A) tails of mRNAs match a short stretch of U residues in the genome which is regarded as the transcription stop/polyadenylation signal or the gene end signal (GE). The start signal is usually separated by several (1-59) residues that are not present in the mRNAs, and these sequences are regarded as intergenic sequences. According to a popular model, the polymerase transcribes the template faithfully until it reaches the oligo-U signal. By a mechanism involving the repeated cycles of backward slippage of the polymerase over the oligo-U stretch leads to the addition of the polyA tail to the mRNA. After these the polymerase scans the intergenic region and resumes transcription at the downstream start sequences (Conzelmann, 1998). The intracistronic sequences and the extracistronic trailer region downstream to the L gene are only copied in replication mode. In addition to this, the temporal order of the transcripts, which also follow the same order, i.e., the leader RNA and the N-mRNA, are the most abundant and the L-mRNA the least (Villareal et. al., 1976, Abraham and Banerjee, 1976). It is still not clear whether this gradient results from a single initiation at the 3' end of the genome RNA followed by a “stop-start” synthesis (Emerson, 1982) or is due to internal initiations coupled with a “cascade” mode of RNA synthesis (Testa et. al., 1980b). The order of the genes, which is highly conserved, is the major determinant of the relative levels of gene expression, since the genes which are close to the single promoter site at the 3' end of the viral genome, are transcribed at higher levels than those that occupy more distal positions. Based on these observations, several scientists have postulated different models for rhabdovirus transcription mechanism.
Model I: The simplest model is the Cleavage Model which states that mRNAs are formed \textit{in vitro} by the cleavage of the growing nascent chain initiated at the 3' end of the genome RNA (Ball \textit{et. al.}, 1976, Banerjee \textit{et. al.}, 1977). Cleavage was proposed to occur at the specific intergenic sites followed by capping and polyadenylation. But the main drawback of this model is the failure to explain the presence of non-equimolar ratios of different mRNAs. Moreover due to the insufficient data of the presence of large uncleaved precursor RNA, this model lacked adequate evidence to persist (Banerjee, 1987).

Model II: The Stop-Start model proposed that the RNA dependent RNA polymerase starts at the 3' end of the genome RNA and synthesizes the leader RNA first, then pauses and the same enzyme complex starts the synthesis of the next mRNA continues up to the 5' end of that gene and adds poly(A) by the oscillation of the U7 sequence present at the intergenic region; the same enzyme complex again initiates synthesis of next mRNA and continues up to the 5' end of the viral mRNA. This phenomenon of restarting the transcription continues up to the 5' end of the L gene. In this journey the complex may fall from the genome RNA but must initiate the synthesis of mRNA from the 3' end of the genome RNA (Emerson, 1982).

This model satisfies the observation that the mRNAs are present in non-equimolar ratio but the RNA polymerase complex must approach the genome at a single site i.e., at 3' end of the genome throughout the transcription process. The experimental support behind this model is that a DI particle, HR-LT2, which contains a leader gene, located approximately 70nt from the 3' end of the genome seems to be obligatory. Another experimental evidence has been presented by (Emerson, 1982) is that in an incomplete reaction mixture containing only ATP and CTP, the reconstituted RNA complex synthesizes only dinucleotide ppAC representing the 5' dinucleotide of the leader RNA.

There are also some observations inconsistent with the above model. This model predicts that the synthesis of leader RNA must be preceded by the synthesis of N mRNA. But in some reaction conditions, e.g., at low salt (Pinney and Emerson, 1982a), in presence of M protein (Pinney and Emerson, 1982b) or in presence of inhibitor aurintricarboxilic acid or vanadylribonuclease complex (Talib and Hearst, 1983) there is preferential synthesis of capped and uncapped oligonucleotides representing the 5' terminus of the N-mRNA. Moreover, synthesis of pppAC may not necessarily represent the initiating sequence of the leader gene. There are domains inside the leader template, where RNA polymerase may initiate and synthesize such dinucleotide. Finally although
this model explains the polarity of transcription, it doesn’t explain the reason for pausing of polymerase at the junction of leader gene and N gene because leader RNA is not polyadenylated. Despite this observations the stop-start model is a highly attractive one because it suggests a similar mechanism of transcription as well as replication, in both modes the polymerase will recognize the same initiation site rather than discriminate between the 3' terminal site and internal polymerase binding sites.

**Model III**: The third one, i.e., multiple initiation model, proposes that the transcription initiation occurs at different internal sites in addition to the leader template (Testa et al., 1980). RNA synthesis thus can start at the putative promoters at the 3' end of any gene but the elongation of each internally elongated RNA depends on the prior transcription of the 3' proximal gene. This pausing of initiated transcript may form the basis of attenuation of the RNA synthesis. This model basically agrees with the stop-start model but differs in that, each mRNA results from a separate initiation site and synthesis of each mRNA is independent. The evidence in support of this model is the isolation of a complex containing leader RNA linked with distinct 5'-terminal of the N-mRNA and the P-mRNA from the transcription reaction (Chanda and Banerjee, 1981). Multiple internal bindings of polymerase also have been demonstrated in reconstitution reaction (Thornton et al., 1984).

### 1.5.2 Replication

For all negative-strand RNA viruses, including the rhabdoviruses, infection begins with the synthesis of viral mRNAs by the viral RNA polymerase of the infecting virus. This RNA synthetic step is known as “Primary Transcription”. The translation of this viral mRNAs leads to the onset of genomic RNA synthesis or replication. Addition of cyclohexamide inhibits viral replication, which indicates that continuous protein synthesis is absolutely needed for viral genome replication (Wertz and Levine, 1973). Unlike the template for primary transcription, which is the negative strand RNP complex, the template for replication is positive strand RNP complex. Since, positive strand genomic RNA is not found in the infected cells, the concomitant binding of the N protein with the growing positive strand genome RNA seems to be the plausible mechanism of positive strand nucleocapsid formation. Until now, two different models have been postulated considering the in-
volvement of N and P protein in this process.

**Model I**: Two main observations: the leader RNA is the first product in transcription and continuous protein synthesis is necessary for viral replication eventually led to the hypothesis of replication (Blumberg *et. al.*, 1981): The viral N protein available in large quantities in the late phase of viral growth, binds with the leader RNA and somehow prevents the termination at the leader RNA junction. The common feature of this model is the speculation that during N protein sufficiency, the nascent RNA transcripts are rapidly encapsidated by a process that releases the viral RNA polymerase from a site-specific attenuation and allows it to synthesize the genomic length RNA. In N protein deficiency, as in *in vitro* transcription, the nascent transcripts are released, new transcripts initiated at specific sites to synthesize the monocistronic mRNAs. Once formed, these mRNAs remain unencapsidated even during N protein sufficiency because they do not contain the nucleation sites, which are absolutely needed for encapsidation. Thus N protein is thought to control the balance between transcriptional and replicative RNA synthesis in infected cells.

**Model II**: Based on the observation that both anti-P antibody and anti-N antibody inhibits viral replication, (Hill and Summers, 1982) have postulated the second model considering the role of both N and P protein in viral replication. Peluso and Moyer (1984) showed the requirement of a putative N-P complex in virus replication *in vitro*. Thus P protein may control the viral replication by making N protein available for encapsidation. In this model, low concentration of N protein can support replication, but at higher concentration, the protein aggregates and this is prevented by the P protein which complexes with N protein (Davis and Wertz, 1982, Patton *et. al.*, 1984).
Chapter One

Introduction

Life cycle of the virus is shown in the figure below:

1.6 N-P interaction in RNA synthesis

The encapsidation of viral genome RNA by N protein to form the nucleocapsids and the association of P protein with these nucleocapsids and free unassembled N protein is very important for viral genome expression. It's known for a long time now that the RNA dependent RNA polymerase, the L protein recognizes the N-RNA template by it's interaction with the template bound P protein (Mellon and Emmerson, 1978). Efficiency of the transcription process depends on the specific
binding of the viral RNA polymerase to the viral nucleocapsids and it has been also seen that independent clusters of P protein bind tightly to the nucleocapsids and these have been demonstrated to activate transcription (Curran, 1996). Thus it’s quite evident that N-P interaction is extremely crucial in the viral life cycle. In the early events of viral genome expression, during primary transcription, the P protein interacts with the N protein in the nucleocapsids and this interaction of the P protein with the assembled N protein is important for effective transcription. In VSV it has been found that C terminal 11 amino acids are responsible for the interaction of the P protein with the N RNA template. From the functional domain studies of the phosphoprotein P it appeared that for transcription the carboxy-terminal is needed in a catalytic amount where as the acidic domain I is needed in a stochiometric amount (Chattopadhyay and Banerjee, 1988). During transcription, the P protein binds with the RNA dependent RNA polymerase L protein and directs it to a promoter site(s) in the N-RNA template to initiate transcription. For efficient transcription to occur the L protein must move along the genomic RNA template. If the L protein has to move along the genomic RNA during transcription then the N protein must be removed from the RNA, which is encapsidates. It’s thought the during primary transcription, the P protein interacts with the N protein in the nucleocapsids and removes the N protein so that the L protein can move along the genome easily. It has been seen that the carboxy-terminal (440-524) amino acids of the N protein is involved in the binding of the P protein (Buchholz, et. al., 1994). It is believed that that the amino terminal acidic domain of the Phosphoprotein P could mimic the negatively charged RNA such that the N protein interacts with it preferentially and leave the genomic RNA uncovered so that the L protein can interact with it (De and Banerjee,1985, Hudson et. al., 1986).

Translation of the primary transcripts results in the intracellular accumulation of viral proteins. After the concentrations of the viral proteins had built up to a certain extent the virus shifts it’s mode from transcription to replication. During the replication process the same viral polymerase ostensibly enters the 3’ end of the virus genome, but now ignores the intergenic signals and reads through the gene junctions (and editing sites) and synthesizes an exact complementary copy of the virus genome and from which subsequently large number of viral genomes are synthesized. Replication and genome assembly are tightly coupled, and it’s believed that this coupling somehow permits the polymerase to ignore the cis acting transcription stop signals. It is believed that the major regulator of the switch is the...
intercellular concentration of unassembled N protein. During replication a soluble form of the N protein interacts with the newly transcribed RNA chains from the RNP complex which in turn leads to successful encapsidation of the genome RNA and thus forming the tightly structured N-RNA complex with the full length positive or negative sense genome RNA inside. Interaction of the soluble form of the N protein with the leader RNA helps in the successful encapsidation of the nascent genomic RNA and helps the viral polymerase to read through the intergenic transcription termination signals. (Blumberg et. al., 1981, Wertz et. al., 1998) But the N protein has a tremendous tendency to form large aggregates, and once aggregated the protein becomes inactive. (Sprague, 1983). The inactive aggregated N protein can bind to the viral leader RNA but cannot encapsidate the genome RNA (Blumberg et. al., 1981). In VSV, it has been found that interaction of N protein with P keeps the N protein in a soluble form in vivo that is capable of enwrapping the de novo synthesized genome RNA (Davis, 1986, Peluso et. al., 1988; Howard, 1989). N-P interaction also may confer specificity towards RNA binding activity of N protein where it channels N pool to viral RNA sequences only (Masters and Banerjee, 1988). So the P protein regulates the presence of replication competent N protein by the formation of N-P complexes. In VSV it has been seen that that the P protein interacts with the N protein with it’s C terminus (Takacs, et. al., 1983) and keeps N in a soluble and active form (Gupta and Banerjee 1997, Howard and Wertz, 1989). In RPV the N terminal 59 amino acids and the C terminal 316-347 amino acids are required for it’s interaction with N protein (Shaji et. al., 1999). In Sendai virus also the N terminal 1-77 amino acids were needed for genome replication and out of that amino acids 33-41 were needed for forming stable complexes with N protein (Curran, et. al., 1995a). In Sendai also the C terminal 29 amino acids were needed for NP interaction (Curran, et. al., 1995a). Similar results were obtained for Measeales virus where the P protein has two independent N protein binding domains in the N terminal and the C terminal half (Harty and Palese, 1995). Similar results were obtained for HPIV39 (Zhao and Banerjee, 1995).

Coimmunoprecipitation experiments, with rabies virus N and P proteins revealed two regions, N terminal 69-138 amino acids and 173-297 amino acids of the P protein were needed for N protein binding. Of these two regions the C terminal amino acids (268-297) was absolutely needed for the N-P interaction (Chenik and Blondell 1998). Later on work done by Tordo et al proved that there are two independent N binding domains in P protein—the weak N terminal domain and a stron-
ger C terminal domain in which the lysine rich motif FSKKYKF is extremely important (Jacob and Tordo, 2001). In VSV infected cells also N-P complexes were found (La Ferla and Peluso, 1989; Masters and Banerjee, 1988, Richardson and Peluso, 1996) and it was proposed that the formation of these complexes regulated viral transcription and replication (Abraham and Banerjee, 1976). Studies on the nucleocapsid protein N has revealed that N protein also has specific domains for it’s interaction with the P protein. In Meseales virus N protein the presence of the amino terminal (4-188) amino acids was needed for the N-P interaction (Bankamp et. al., 1996). In VSV it was found that deletion of C terminal 10 amino acids reduced the strength of NP interaction (Takacs et. al., 1983). For Sendai virus also deletion of C terminal 47 amino acids was deleterious for N-P interaction (Homann et. al., 1991). In RPV however both the C terminal and the N terminal end of the N protein was needed for N-P interaction (Shaji et. al., 1999). In Sendai also other than the C terminal 47 amino acids the N terminal half is also needed for the N-P interaction (Homann et. al., 1991). Studies with the N protein of negative sense RNA viruses suggests that after forming the assembled nucleocapsid structure about 80% of the N terminal half of the N protein remains hidden and only C terminal half is exposed, (Heggness et. al., 1981, Gill et. al., 1988, Buckland et. al., 1989, Ryan et. al., 1993, Dietzschold et. al., 1987, Kouznetzoff et. al., 1998 Despande and Portner 1984, Buchholz et. al, 1993, (Schoehn, 2001)). So it is conceivable that prior to the encapsidation process the amino terminal end is more accessible for N-P interaction. In Sendai (Homann et. al., 1991) Meseales (Bankamp et. al., 1996) Bovine respiratory syncytial virus (Krishnamurty et. al., 1998) N terminal half of N protein is needed for it’s interaction with P protein. But after encapsidation the exposed C terminal end becomes important for the N-P interaction. In Sendai (Buchholz et. al., 1994), Rabies (Schoehn, 2001) and RPV (Shaji, 1999) the extreme C terminal end of N protein is needed for the N-P interaction. So this N-P interaction through the various specific domains in P and N protein not only helps in maintaining the N protein in a replication competent and soluble form but it also helps the N protein to interact with the virus specific leader RNA. Based on all these information it can be suggested that the N-P interaction is extremely important for the viral life cycle and thus makes it an important aspect of study.