Review of Literature
2.1 Japanese Encephalitis

Japanese encephalitis (JE) is the single most important cause of acute encephalopathy in human beings. JE is widely distributed in Asia, where up to 50,000 cases are estimated to occur annually, of which around 20% result in fatality, though the disease is greatly underreported (Tiroumourougane et al. 2002; Umenai et al. 1985). Though a disease resembling JE was recognized as early as 1871, a filterable agent was extracted from human brain from a severe epidemic case in Japan in 1924. The agent was, however, not identified or characterized. The virus responsible for JE was recovered from a sick horse's brain in 1935 (Miyake 1964) and was serologically established as the prototype Nakayama strain. In India, JE was clinically diagnosed for the first time in 1955 at Vellore, Tamilnadu. Since then, outbreaks have been reported from almost 25 states and union territories of India. The occurrence of JE cases and deaths has shown a rising trend in the recent years. At present, JE is not only endemic in many areas; it is also spreading to the naïve non-endemic areas, probably due to ecological changes and extensive traveling of people. Thus, epidemics and sporadic cases of JE can be found throughout the temperate and tropical zones of Asia including India.

2.1.1 Japanese Encephalitis: Clinical features and Pathology

The virus responsible for JE infection is a flavivirus called Japanese encephalitis virus (JEV) which is transmitted to humans by the culicine mosquitoes (Fig. 2.1). The virus is transmitted in a zoonotic cycle among mosquitoes and vertebrate-amplifying hosts, chiefly pigs and wading birds (Banerjee 1996). Birds and pigs are effective viremic amplifying hosts, which serve as the source of infection of mosquito vectors (Scherer and Buescher 1959). Pigs are known to be the amplifiers of JEV because of the prolonged and high titers of viremia without clinical symptoms. Humans are considered as the dead end hosts, as the brief periods of viremia and low titers of virus do not facilitate transmission. In temperate regions, virus transmission is detectable in July in mosquitoes, birds and pigs and weeks after in humans.

Following an infective mosquito bite, the JE incubation period is 6-16 days that is followed by the rapid onset of the disease. The initial viral multiplication may take place in local regional lymph nodes. The first 2-4 days constitute the prodromal phase.
Fig. 2.1: Transmission cycle of Japanese encephalitis virus.

Pigs are the primary host for JEV which gets transmitted to humans by mosquitoes feeding on viraemic pigs. Humans are dead end hosts and do not participate in perpetuation of virus transmission. Vector species vary, but culicine mosquitoes are responsible for the transmission of the virus to humans.
Review of Literature

comprising of headache, fever, chills, anorexia, nausea, vomiting, dizziness and drowsiness. In children, abdominal pain and diarrhea may be prominent. These symptoms are followed by appearance of nuchal rigidity, photophobia, altered states of consciousness, hyper excitability and varying objective neurological signs including dull, mask like facies, muscular rigidity, cranial nerve palsies, tremulous eye movements, coarse tremors of extremities, involuntary movements, generalized and localized paresis in coordination of reflexes. Sensory deficits are rare. Paralysis of upper extremities is more common than paralysis of legs. Convulsions are frequent in children but occur in less than 10% of adult patients. Death occurs on 5-9th day. The peripheral white blood cell (WBC) count is only mildly elevated but urinary tract symptoms are common during acute phase of illness. The electro encephalogram (EEG) is abnormal with decreased electrical activity and dysrhythmia. Approximately 5-40% cases result in fatality, though in case of epidemics, the rates can be as high as 70%. Around 45-70% of survivors have neuropsychiatric sequelae, which are particularly severe in children (Kumar et al. 1993) that include Parkinsonism, convulsive disorders, paralysis, mental and psychiatric disorders.

Following the initial replication at the site of entry, JEV is transported across the cerebral blood vessels. Though the pathway used by the virus to cross the blood brain barrier is not well defined, electron microscopic studies in JEV infected mice show that the virion gets attached to endothelial cells and gets transported across blood vessels through an endocytic vesicle, where it is taken up by a perivascular cell and transported into the Central Nervous System (CNS) (Liou and Hsu, 1998). During the acute stage of the infection, congestion, edema and small hemorrhages are found in the brain. Microscopic lesions include neuronal degeneration and necrosis, neuronophagia, microglial proliferation forming glial nodules and perivascular inflammation.

2.1.2 Japanese Encephalitis: Diagnosis and Treatment

In the fatal cases, JEV antigen can be isolated from brain (Mathur et al. 1990). The virus cannot be isolated from blood but can be isolated from cerebrospinal fluid (CSF) in upto one third of patients during acute phase. The diagnosis can also be done by immunofluorescence examination of the cerebrospinal fluid for the presence of the viral
antigen. An IgM capture ELISA is widely used on CSF for the rapid diagnosis of JE infection. There is no specific treatment available against JE. Good supportive patient care is essential for the favorable outcome of the disease.

2.2 Japanese Encephalitis Virus

JEV is a member of the flavivirus genus of *Flaviviridae* family of positive strand RNA viruses. YF virus is the prototype member of the flavivirus genus that derives its name from the Latin word *flavus* meaning yellow. The other genera within the *Flaviviridae* family include the pestiviruses (from the Latin word *pestis*, meaning plague) and the hepaciviruses (from the Greek word *hepatos*, meaning liver) (Sumiyoshi et al. 1987; Westaway et al. 1985). Bovine Viral Diarrhea virus (BVDV) is the prototype pestivirus whereas Hepatitis C virus (HCV) is the prototype hepacivirus. In addition to these genera, a group of unassigned viruses, the GB agents, are awaiting formal classification within the family. Though the members of the *Flaviviridae* family exhibit diverse cross reactivity, they share similarities in virion morphology, genome organization and presumed RNA replication strategy.

The genus flavivirus is currently composed of around 70 viruses, which are further subdivided into antigenic complexes according to serological criteria, or into clusters, clades and species on the basis of molecular phylogenetics (Kuno et al. 1998). The flavivirus genus, apart from JEV and YFV, includes such medically important viruses as the West Nile virus (WNV), Tick-Borne Encephalitis virus (TBEV), and all four of Dengue viruses (DEN). Flaviviruses are mainly divided into three clusters, namely mosquito-borne (WNV, JEV, Kunjin, DEN, YFV etc., 34 viruses), tick-borne (TBE etc., 17 viruses) and those with no known vectors (22 viruses). JEV is the prototype of the JE antigenic complex (Poidinger et al. 1996) that consists of WNV, DEN, St. Louis Encephalitis virus (SLE) and Murray Valley Encephalitis virus (MVE).

YFV, DEN, and in the recent years WNV, are the most widely studied flaviviruses. Our knowledge about JEV is patchy and altogether absent for certain features. Thus, information on JEV is based on the extrapolation of our knowledge on the closely related flaviviruses such as YFV and WNV. A review of the information available
for different flaviviruses is presented below together with the specific information on JEV, whereever available.

2.2 The Virion

The flavivirus particles are enveloped spheres of ~ 50 nm diameter surrounded by the lipid bilayer. Owing to the lipid envelope surrounding the virions, flaviviruses are readily inactivated by organic solvents and detergents. There are three viral proteins associated with the virions; these are envelope (E), membrane (M) and capsid (C) proteins. The E protein is the major surface protein of the viral particle and interacts with receptors on cell to mediate virus-cell membrane fusion. Antibodies that neutralize virus infectivity usually recognize the E protein and mutations in the protein can affect the virus virulence (Heinz 1986). The M protein is a small proteolytic fragment of the precursor prM protein. Removal of the lipid envelope with non-ionic detergents reveals discrete nucleocapsids, which consist of C protein and the genomic RNA (Russell 1980). In addition to mature virions, smaller non-infectious particles are released from flavivirus-infected cells. These particles are called slowly sedimenting hemagglutinins (SHA) because, like virions, they can agglutinate red blood cells at low pH. SHA are smaller in size than virions and consist of E and M proteins, but lack nucleocapsids (Smith et al. 1970).

2.2.2 The Genome

The genome of flaviviruses is a positive-sense, single-stranded RNA of about 11 kilo bases in length with a type 1 cap (m7G5’ppp5’A) at the 5’- end but no polyadenylate (polyA) tail (Cleaves and Dubin 1979; Wengler et al. 1978). The genome encodes a single open reading frame (ORF), flanked by the 5’- and 3’-noncoding regions (NCRs) of ~ 100 and 400-700 nucleotides, respectively (Fig. 2.2). These regions contain sequences and RNA structures that are likely to serve as cis-acting elements directing the processes of genome amplification, translocation and/or virus maturation. The genome of flaviviruses, as for all positive-sense RNA viruses, is infectious (Peleg 1969), that is, as soon as genome gets uncoated inside the host cell, it can serve as the template for the synthesis of the viral proteins.
Fig. 2.2: Flavivirus genome organization.

(A) RNA structure of flavivirus genome is illustrated. The genome is approximately 11 kb long, containing approximately 100 nucleotides long 5'-NCR, 400-700 nucleotide long 3'-NCR and a 5'cap. Initial one-third region of the single ORF codes for structural protein and the rest codes for non-structural proteins of the virus.

(B) The protein expression and polyprotein processing strategy of flaviviruses is illustrated. The polyprotein is cleaved by various viral or host proteases to produce three structural and seven non-structural proteins.
The ORF, with the help of the cellular machinery, gets translated into a large polyprotein of ~3400 amino acids, which is then co- and post-translationally cleaved into at least 10 proteins. The sequence of proteins that are encoded by the genome is as follows. The N-terminal quarter of the polyprotein codes for structural proteins (C-prM-E) and the remainder codes for seven non-structural proteins in the order NS1-NS2A-NS2B-NS3-NS4A-NS4B-NS5 (Rice et al. 1985). The host signal peptidase is responsible for cleavage between C-prM, prM-E, E-NS1 and near the C-terminus of NS4A. A virus-encoded serine protease is responsible for cleavage between NS2A-NS2B, NS2B-NS3, NS3-NS4A, NS4A-NS4B, and NS4B-NS5 proteins. There is an unknown enzyme responsible for the cleavage of the NS1-2A proteins.

The JEV genome is 10976 bases long for the majority of the strains (Chambers et al. 1990a; Vrati et al. 1999). It differs by a few nucleotides in some strains. For example the P20778 and the Ling isolates are 10977 and 10951 bases long, respectively. In the case of JEV, the 5' and 3' NCR length is 98 and 585 bases, respectively (Sumiyoshi et al. 1987; Vrati et al. 1999). The ORF encodes for a polyprotein of 3432 amino acids which is cleaved into three structural and seven non-structural proteins.

2.2.3 The Structural Proteins

**Capsid protein (C):** The capsid is a highly basic protein of ~11 kDa, with charged residues clustered around N- and C-terminal separated by a short hydrophobic region that mediates the membrane association (Boege et al. 1983; Khromykh and Westaway 1996; Rice et al. 1985).

**Membrane protein (M):** The glycoprotein precursor of the M protein, prM of ~26 kDa, is translocated into endoplasmic reticulum (ER) by the C-terminal hydrophobic domains or signal sequence. Here, the prM protein is processed and cleaved by the host signal peptidase into M protein. The conversion of immature virus particles to mature virions occurs in the secretary pathway and coincides with the cleavage of prM to M (Stadler et al. 1997). The pr segment of the prM protein is thought to stabilize the E protein in the reduced pH environment of early secretary pathway (Guirakhoo et al. 1991; Guirakho et al. 1992). Following the cleavage, pr segment is secreted and M associates with the mature virions (Murray et al. 1993).
**Envelope Protein (E):** The E protein is the major surface protein of mature virions with molecular mass of ~58 kDa. It contains 12 highly conserved cysteine residues that form intramolecular disulfide bonds and is often glycosylated (Chambers et al. 1990b; Kanaya et al. 1997; Nowak and Wengler 1987). The protein plays an important role in binding and membrane fusion of virion during entry and assembly of mature virion. This protein is the major target for the neutralizing antibodies against virus.

### 2.2.4 The Non-Structural Proteins

**NS1:** The NS1 is a ~46 kDa protein that is retained mostly in the cell but is also found on the cell surface and is secreted slowly from the virus-infected mammalian cells (Mason 1989; Post et al. 1991; Smith et al. 1970; Winkler et al. 1988). It is translocated into the ER and released from the E protein by host signal peptidase (Chambers et al. 1990b; Falgout and Markoff 1995). Around 30 minutes after synthesis, NS1 forms homodimers that are highly stable and resistant to denaturation (Winkler et al. 1989). Due to dimerization, it acquires a partial hydrophobic character and hence associates peripherally with membrane. NS1 co-localizes with vesicle packets (VPs), the most likely sites of RNA replication (Mackenzie et al. 1996; Westaway et al. 1997). Though the role of this protein is not clear in RNA replication, mutations in the glycosylation sites of this protein lead to a dramatic effect on RNA replication (Muylaert et al. 1996) and virus production (Pryor et al. 1998).

**NS2A:** The NS2A is a hydrophobic protein of relatively small size of ~22 kDa. The N-terminus of NS2A is generated via the cleavage of NS1-NS2A by an unknown ER-resident host enzyme (Falgout and Markoff 1995), whereas C-terminus is generated by the viral serine protease cleavage in the cytoplasm. NS2A was localized by cryo immunogold staining to VPs, the presumed sites of RNA replication (Mackenzie et al. 1998). The role of this protein has been predicted in coordinating the shift between RNA packaging and RNA replication (Khromykh et al. 2001a).

**NS2B:** The NS2B is a small, membrane-associated protein of ~14 kDa molecular mass. The protein forms a complex with NS3 and is a necessary cofactor for the serine protease activity of NS3 (Arias et al. 1993; Chambers et al. 1991; Chambers et al. 1993; Jan et al. 1995; Yusof et al. 2000). The cofactor activity of NS2B is encoded in a
conserved hydrophilic region of ~40 amino acid residues, flanked by hydrophobic regions that mediate membrane association (Clum et al. 1997).

**NS3**: The NS3 is the second largest viral protein of ~70 kDa molecular mass. This protein is well conserved among flaviviruses (Mandl et al. 1989) and is a multifunctional protein containing several enzyme activities that are involved in polyprotein processing and RNA replication. This protein is associated with membrane by the interaction with NS2B (Chambers et al. 1993; Clum et al. 1997) and has been co-localized to VPs. The N-terminal one-third of NS3 is a serine protease. The protease preferentially cleaves after the adjacent basic residues and before an amino acid containing a small unbranched side chain (Chambers et al. 1993; Chambers et al. 1995; Nestorowicz et al. 1994) and can be inhibited by aprotinin and substrate analogues (Leung et al. 2001; Murthy et al. 2000). The C-terminal region has been implicated in RNA replication, encoding regions of significant homology to RNA helicases (Gorbalenya et al. 1989). RNA stimulated NTPase activity has been demonstrated for several full-length and C-terminal NS3 proteins (Kuo et al. 1996; Li et al. 1999; Takegami et al. 1995). The precise role of helicase melting activity in RNA replication remains unknown though functions have been predicted in melting RNA secondary structures and thus template recognition, increasing the polymerase processivity by eliminating secondary structures, or resolving duplexes formed during the process of replication. In this regard NS3 was shown to bind the 3'-stem loop in the 3'-NCR in association with NS5 and that NS3 NTPase activity was enhanced in the presence of NS5 (Chen et al. 1997; Cui et al. 1998). In addition to the NTPase activity, the C-terminal region of NS3 encodes an RNA triphosphatase (RTPase) activity (Wengler and Wengler 1993), though it is unclear whether this activity is different from NTPase activity.

**NS4A and NS4B**: NS4A and NS4B are very small, hydrophobic proteins of ~16 and 27 kDa molecular masses, respectively. The function of these proteins are not known but a role has been suggested in RNA replication (Lindenbach and Rice 1999), which is supported by the co-localization of these proteins to VPs as well as to convoluted membranes (CM), which are thought to be involved in polyprotein processing.
NS5: NS5 is the most conserved, largest viral multifunctional protein of ~103 kDa molecular mass. It contains N-terminal RNA cap processing activity and C-terminal RNA-dependent RNA polymerase (RdRp) activity. The N-terminal region of NS5 was found to contain homology with S-adenosyl-methionine (SAM)-dependent methyltransferases suggesting that this protein is involved in the formation of the 5' cap (Egloff et al. 2002; Koonin 1993). The C-terminal of NS5 contains significant homology to RdRp of other positive-strand RNA viruses (Koonin 1993; Rice et al. 1985) and the polymerase activity has been confirmed with recombinant NS5 (Ackermann and Padmanabhan 2001; Guyatt et al. 2001; Tan et al. 1996). It has been demonstrated that NS5 forms a complex with NS3 (Kapoor et al. 1995) and can stimulate NS3 NTPase activity. Both of these proteins have been shown to interact with 3'-stem loop of the 3'-NCR (Chen et al. 1997).

2.2.5 The Noncoding Regions (NCRs)

There are two NCRs in flavivirus genome flanking the ORF. The 5'-NCR of flaviviruses is about 100 nucleotides long and the 3'-NCR is about 400 to 800 nucleotides long. The NCRs may be involved in the initiation of the negative and positive strand synthesis, in processes related to switching from negative strand to positive strand synthesis and vice versa, and regulation of translation and packaging of genome into virions. Specific features of the NCRs are given below.

5'-NCR: The sequence of the 5'-NCR is not well conserved among different flaviviruses and only short regions of sequence are conserved. But there is significant sequence homology in the 5'-NCR sequence among viruses of the same subgroup and an almost complete conservation of the nucleotide sequence among different strains of the same virus (Brinton and Dispoto 1988; Cahour et al. 1995; Hahn et al. 1987). The 5'-terminal two nucleotides of the plus strand sequence (5'-AG) complementary to the 3'-terminal two nucleotides (UC-3') are completely conserved among mosquito- and tick-borne viruses (Wengler and Wengler 1981). Apart from this, the most conserved linear sequence feature near the flavivirus 5'-NCR is the cyclization sequence (5'-CS) (Hahn et al. 1987; Khromykh et al. 2001; Mandl et al. 1993). The 5'-CS actually lies within the ORF, about 30-40 nucleotides downstream of the start codon (Fig. 2.3). A complementary
Fig. 2.3: Features of the 5'- and 3'-NCRs of mosquito-borne flavivirus genomes.

Horizontal lines represent nucleotide sequences of indicated flavivirus genome NCRs. Upward vertical ticks in the 5'-NCR indicate the location of the 5' most start codon for translation of the ORF in the flavivirus genome. Downward vertical ticks indicate the location of translational stop codon at the 5' end of the 3'-NCRs. The flavivirus-conserved stem-loop (SL) secondary structure predicted to form at the 3' end of all genomes is indicated by a loop at the end of the linear sequence. The relative locations of the conserved eight nucleotide cyclization sequences, 5'CS and CYC, respectively, at the 5' and 3' ends of the genome are indicated by open boxes with cross-hatching. The relative location of the conserved sequence, CS2, in all genomes is indicated by a stippled box, and the location of the CS2 tandem repeat sequence, RCS2, in all but the Yellow fever virus genome is indicated by a second stippled box upstream of CS2. Similarly, the relative locations of the conserved tandem repeat sequences, CS3 and RCS3, are indicated by black, cross-hatched boxes in genomes of viruses of the JE subgroup. The location of the three tandem repeat sequences in the upstream portion of the Yellow fever virus genome 3'-NCR are indicated by open rectangles.
sequence to 5'-CS exists in 3'-NCR and hence the motif is named cyclization sequence, as in theory base pairing between 5'- and 3' cyclization sequences would permit the formation of circularized plus strand during initial phases of RNA replication, which could be stabilized by various host/viral proteins. For the mosquito-borne viruses, there is only one cyclization sequence whereas for tick-borne viruses, two potential cyclization sequences have been identified (Fig. 2.4), designated as C1 and C2 (Khromykh et al. 2001; Mandl et al. 1993). Tick-borne virus 5'-NCR contains a very short ORF at the very 5'-end of the genome. The first cyclization sequence lies downstream from this mini ORF and wholly lies within the 5'-NCR. The C2 lies downstream of C1, within the true ORF, in a position analogous to that of the 5'-CS in mosquito-borne virus genomes (Khromykh et al. 2001). The importance and requirement of the base pairing of the cyclization sequence for RNA replication has been demonstrated. Mutational analysis in full length infectious clone of DENV4 demonstrated that the cyclization sequence was indispensable for virus replication (Men et al. 1996). Mutation of either 5' or 3' cyclization sequence such that base pairing between them was abrogated resulted in total loss of in vitro RdRp activity but when both the sequences were mutagenized to complementary sequences, RdRp activity was again detected (Brinton and Dispoto 1988).

Despite the differences between sequences of 5'-NCRs of different flaviviruses, secondary structure of similar size, shape and predicted thermodynamic stability could be formed. This structure consisted of a stem with a small top loop and a larger side loop. In most cases, a second short stem could be formed if some bases downstream to start codon were included.

3'-NCR: The 3'-NCRs of flavivirus genomes exhibit great variability, although several conserved features and secondary structures have been elucidated and predicted. As stated above, the 3'- terminal two nucleotides (UC-3') complementary to 5'-AG in the 5'-NCR are highly conserved among flaviviruses. Similarly, the 3'-CS complementary to 5'-CS is well conserved and the locus of the sequence in mosquito-borne viruses is just upstream to the conserved terminal 3'-stem loop (3'-SL) structure (described later). For tick-borne viruses, C1' sequence (complementary to C1) is localized to the most upstream region of 3'-SL and C2' sequence is located further upstream from C1'. Several non-conserved complementary bases are present upstream from the 5'-CS and downstream
Fig. 2.4: Features of 5’- and 3’-NCRs of tick-borne flavivirus genomes.

Horizontal lines represent nucleotide sequences of indicated flavivirus genome NCRs. Upward vertical ticks numbered one in the 5’-NCR indicate the location of the 5’ most start codon of a very short ORF; the start codon for translation of the long ORF in the TBE genomes is represented by an upward tick numbered two. Stop codon for the short ORF in the 5’-NCR and for the long ORF at the 5’ end of the 3’-NCR are indicated by downward vertical ticks. Four of six different possible nucleotide sequence motifs for 3’-NCRs of TBE virus are indicated on the right as A-D. The flavivirus-conserved stem-loop (SL) secondary structure predicted to form at the 3’ end of all genomes is indicated by a loop at the end of the linear sequence. Black boxes indicate the relative locations of the cyclization sequences, C1 and C2 in the 5’-NCR and C1’ and C2’ in the 3’-NCR. The relative location of TBE virus-conserved repeat sequences R1, R1’, R2, and R3 in the 3’-NCR are indicated by cross-hatched or stippled boxes.
from 3'-CS. The base pairing of cyclization sequence could be further stabilized by these complementary sequences (Hahn et al. 1987).

The 3'-NCR of the mosquito-borne viruses contains a conserved pentanucleotide sequence 5'-CACAG, which forms part of an unpaired region in the loop contained within the 3'-SL. It has been suggested that the pentanucleotide may play a role in binding of cellular and/or viral proteins to the conserved 3'-SL during RNA replication (Khromykh et al. 2001).

The conserved tandem repeat sequences CS2 and RCS2 were initially described by (Hahn et al. 1987). These sequences are present in mosquito-borne viruses and in tick-borne viruses upstream of CS1. The JE subgroup viruses have an additional pair of repeat upstream of RCS2 in 3'-NCR termed as CS3 and RCS3. Very little is known about the function of these conserved tandem repeats in mosquito-borne flaviviruses. The DENV4 genome could tolerate the deletion of more than half of its 3'-NCR including tandem repeats, provided the deleted sequence were upstream of nucleotide -113, without complete loss of viability (Men et al. 1996). The downstream of -113 nucleotides form stable 3'-SL structure that is well conserved among flaviviruses.

The 3'-SL, a stable stem-loop secondary structure of same thermodynamic stability, has been proposed for the 3' termini of all flavivirus genomes (Brinton et al. 1986; Hahn et al. 1987; Irie et al. 1989; Rice et al. 1985; Wengler and Castle 1986). The 3'-SL has a predicted thermodynamic stability of -40 to -45.2 kcal, largely due to the free energy state of the long stem region which consisted of about 30 hydrogen bonded base pairs. The nucleotide sequence of the 3'-SL is only partially conserved among unrelated flavivirus species. The 3'-SL nucleotide sequence is best conserved among the closely related members of the same subgroup or clade, e.g. among species assigned to JE subgroup. Brinton et al. (1986) provided the physical and biochemical evidence for the existence of the 3'-SL. A tertiary pseudoknot structure is formed by further interaction among the nucleotides of the 3'-SL (Shi et al. 1996). This is formed by hydrogen bonding the nucleotides of one strand of the long stem and nucleotides in the loop structure atop an adjacent short stem upstream of the 3'-SL. Such atypical hydrogen bonding has been described for other RNA species as well.
Three regions of secondary structure could be defined in the 3'-NCR sequence based on the analysis of published flavivirus 3'-NCR sequence using a generic algorithm that purportedly simulated the natural folding pathway taking place during RNA elongation and allows prediction of tertiary interaction as well as the secondary structure (Proutski et al. 1997). Region I consisted of nucleotides upstream of CS1, CS2 and RCS2 (Fig. 2.5). This region in all flaviviruses contains high variability in linear nucleotide sequence but formed very similar long hairpin structure with a branching stem loop side structure, or in some cases a bulge loop on the 5' side of the main hairpin. Nucleotides in region II included those assigned to CS2 and RCS2 plus intervening sequences. Unlike region I, region II nucleotides exhibited differences in folding pattern that divided viruses into three groups: (a) DEN and JE subgroup, (b) tick-borne viruses, and (c) YFV. JE subgroup was shown to contain tandem repeat sequences that potentially formed conserved short stem loop structures. Region III consisted of nucleotides traditionally believed to constitute CS1 and the 3'-SL.

### 2.3 Flaviviruses replication

The flaviviruses replicate efficiently in almost all cell types of mammalian and mosquito origins. The replication cycle can be divided into a number of steps, details of which are presented below (Fig. 2.6).

**Entry:** Flaviviruses attach to the host cell surface via an interaction of the E protein (He et al. 1995; Hung et al. 1999) with one or more cellular receptors and these receptors vary among different cell types. Several host surface proteins have been described as putative receptors by virtue of their binding to virus neutralizing anti-E antibodies (Bielefeldt-Ohmann 1998; Kimura et al. 1994; Kopecky et al. 1999; Marianneau et al. 1996; Munoz et al. 1998; Ramos-Castaneda et al. 1997; Salas-Benito and del Angel 1997). For example, two cell surface proteins of 40- and 45-kDa size have been suggested to be putative receptors of DENV in C6/36 cell lines (Salas-Benito and del Angel 1997) while in human and mouse neuroblastoma cell lines, a 65-kDa trypsin-sensitive cell membrane protein has been predicted to be the receptor for DENV (Ramos-Castaneda et al. 1997). Similarly, during a study on susceptibility of various cell lines to infection of JEV, it was seen that highly susceptible Vero cell line binds JEV specifically.
Fig. 2.5: Predicted secondary structure of the 3′-NCR of JE serogroup flaviviruses.

The sequence used for construction of the secondary structure model is from the JaOArS982 strain of JE virus (GenBank accession no. M18370), although very similar structures are found in all JE serogroup viruses. The conserved sequence motifs (CS2, RCS2 and TL2) are shown in bold (Proutski et al., 1997).
Various steps involved in the life cycle of a typical flavivirus are illustrated. Binding and uptake are believed to involve receptor-mediated endocytosis via cellular receptors specific for viral envelope proteins. Fusion of the virion envelope with cellular membranes delivers the nucleocapsid to the cytoplasm, where translation of the genome RNA occurs. All known viral proteins are produced as part of a single long polyprotein of more than 3,000 amino acids that is cleaved by a combination of host and viral proteases. RNA replication occurs in cytoplasmic replication complexes that are associated with perinuclear membranes, and it occurs via synthesis of a genome-length minus-strand RNA intermediate. Progeny virions assemble by budding through intracellular membranes into cytoplasmic vesicles. These vesicles follow the host secretary pathway, fuse with the plasma membrane, and release mature virions into the extracellular compartment.

Fig. 2.6: Schematics of the flavivirus life cycle.
than a low-susceptibility NRK cell line. Upon further analysis it was found that JEV binds to a 74K molecule, which was present in membrane fractions of Vero cells but not in that of NRK cells (Kimura et al. 1994). Another mechanism of flavivirus binding called the antibody-dependent enhancement (ADE) has also been demonstrated for various flaviviruses (Peiris and Porterfield 1979; Phillpotts et al. 1985; Schlesinger and Brandriss 1983). In this mechanism, opsonization with flavivirus- reactive immunoglobulins enhances virus particle binding and infection of cells expressing immunoglobulin Fe receptors. After binding, virions are internalized into clathrin-coated pits via receptor-mediated endocytosis (Gollins and Porterfield 1985; Gollins and Porterfield 1986; Ishak et al. 1988; Nawa 1998; Ng and Lau 1988). Virions are later found in pre-lysosomal endocytic compartment where low pH induces fusion between the virus and host cell membranes to release the virus nucleocapsid into the cytoplasm (Gollins and Porterfield 1985; Gollins and Porterfield 1986; Heinz et al. 1994; Nawa 1998). The viral genome is released into the host cytoplasm by the nucleocapsid disassembly, which is not yet fully understood.

**Virus Genome Translation and Replication:** The plus-sense flavivirus genome acts as the mRNA and is translated as a large polyprotein that is processed co- and post-translationally by cellular and viral protease into three structural (C, M and E) and seven non structural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5). Flavivirus infection induces rearrangement of cytoplasmic membrane in the peri-nuclear regions. The earliest of these events is the proliferation of the ER (Murphy et al. 1968; Ng and Hong 1989), followed by the appearance of smooth membrane structure around the time of early logarithmic virus production. Smooth membrane vesicles appear as clustered vesicles containing electron-dense material within the lumen of ER (Cardiff et al. 1973; Heinz et al. 1994; Leary and Blair 1980; Murphy et al. 1968; Ng et al. 1994b). In samples prepared via cryo-sectioning, these structures appear as VPs, clusters of vesicles, bound by a smooth outer membrane (Mackenzie et al. 1999). During later times of infection, CMs are found adjacent to VPs. CM appear as randomly folded or ordered membranes (sometime referred to as paracrystalline arrays) that are contiguous with the ER (Leary and Blair 1980; Murphy et al. 1968). Based on the localization of viral RNAs and replication
proteins, it appears that VPs are sites of RNA replication (Mackenzie et al. 1996; Mackenzie et al. 1998; Westaway et al. 1997; Westaway et al. 1999).

The viral RNA replicase is assembled from the nonstructural proteins, together with the genomic RNA template and presumably some host factors, on the induced cytoplasmic membranes, probably through interaction involving the nonstructural proteins. Replication begins with the synthesis of a negative strand RNA, which serve as a template for the synthesis of additional positive strand genomic RNA. The viral RNA synthesis is asymmetric, leading to a 10- to 100-fold excess of the positive strands over the negative strands (Cleaves et al. 1981; Muylaert et al. 1996). The negative strands continue to accumulate throughout the infection and have been isolated exclusively in the double stranded form (Cleaves et al. 1981; Wengler et al. 1978). Three major species of flavivirus RNAs have been described; one type sedimenting at 40S, another at 20S and a heterogeneous population at 20S to 28S (Cleaves et al. 1981; Wengler et al. 1978). The 40S RNA is insoluble in 2M LiCl, fully sensitive to RNase treatment and is indistinguishable from genomic RNA (Wengler et al. 1978). The 20S RNA frequently termed as the replicative form (RF) is soluble in 2M LiCl and is resistant to single strand specific RNase, is a double-stranded RNA species and is likely to be a transiently stable duplex of viral positive and negative strand RNAs (Baltimore 1968; Chu and Westaway 1985; Cleaves et al. 1981; Wengler et al. 1978). The heterogeneous (20S to 28S) RNA is insoluble in 2M LiCl and is partially resistant to RNase indicating that it contains both single-stranded and double-stranded regions and are described as replicative intermediate (RI) RNAs that most likely contain duplex regions and newly synthesized positive strand RNA displaced by the nascent strands undergoing elongation (Chu and Westaway 1985; Cleaves et al. 1981). Pulse chase analysis has indicated that RF and RI RNAs are precursors to 40S (genomic) RNA (Chu and Westaway 1985; Cleaves et al. 1981). This mode of replication can be described as semi conservative and asymmetric.

The RdRp activity has been studied in vitro using crude infected cell extracts and all three RNA forms (Chu and Westaway 1985; Grun and Brinton 1986; Grun and Brinton 1987; Grun and Brinton 1988; You and Padmanabhan 1999). However, these in vitro systems appear to involve chain elongation rather than de novo initiation, and the complete process of flavivirus replication has not yet been fully reconstituted in an in vitro model.
Virion Assembly and Release: Ultrastructural studies have indicated that virion morphogenesis occurs in association with intracellular vesicles. Mature virions are thought to assemble by budding into the ER (Deubel et al. 1981; Hase et al. 1987a; Hase et al. 1987b; Ishak et al. 1988; Ko et al. 1979; Leary and Blair 1980; Mackenzie and Westaway 2001; Matsumura et al. 1977; Murphy et al. 1968; Sriurairatna et al. 1973; Sriurairatna and Bhamarapravati 1977; Wang et al. 1998). Budding intermediates and clearly distinguishable nucleocapsids have not been frequently observed, suggesting the process of assembly to be rapid and coupled to RNA replication (Khromykh et al. 2001a). Nascent virions are believed to be transported by bulk flow through the secretory pathway to the cell surface, where exocytosis occurs (Chambers et al. 1990b; Heinz et al. 1994; Mason 1989; Nowak et al. 1989). Budding of virions at the plasma membranes has been described occasionally but this does not appear to be the major mechanism for virion release (Hase et al. 1987b; Ng et al. 1994a; Ng et al. 2001; Ohyama et al. 1977; Smith et al. 1970; Sriurairatna and Bhamarapravati 1977).

2.4 RNA virus replication: role of host cell proteins

Viruses are obligate intracellular parasites that depend on their host cells for replication. We know a great deal about DNA replication that requires a multitude of proteins besides the DNA polymerase. The RNA replication is, therefore, likely to require several proteins besides the RdRp encoded by the viral genome. Most of the RNA viruses have small size genome that codes for only a few proteins. It is thus possible that some of the proteins involved in RNA replication are of the host origin. However, there is no RNA replication in eukaryotic cells and hence no machinery for this process is needed in these cells. It could, therefore, be argued that RNA viruses make use of the host cell proteins by subverting their normal function in the host. There are several other issues that suggest a role of host cell proteins in RNA virus replication. These are listed below.

1. Most RNA viruses are host and tissue specific, meaning that they will grow in some cells while not in others. Also, the rate at which virus multiply in different cells varies. This may be related to the presence or absence of a certain protein or its amounts in the host cell.
2. The sequence as well as structure of the 3'-ends of both the plus and the minus strands is totally different but the RdRp must recognize both of them to start synthesis of the corresponding complementary strands. Besides, the plus strand is synthesized \(~10-100\) folds more than the minus strand. This process may make use of different host proteins to constitute replicase complexes of different specificity and different affinity.

3. The viral genome is released in the cytoplasm but then reaches the intracellular membranes where the replication complex has to be assembled. Host cell proteins may be required for this RNA transportation process.

The conserved structures and sequences in the 5'- and 3'-NCRs despite sequence diversity suggest their possible functional relevance in viral replication cycle. Thus, the NCRs might play role in minus strand synthesis, switch from minus to plus strand synthesis, initiation of translation of genome, and packaging of genome into the core of the virus during assembly. There is increasing evidence that circularization of mRNA is an important step in protein synthesis, presumably because it allows ribosomes to be redirected from the 3'- to the 5' end of the mRNA for a new round of translation and probably because it ensures that only the intact mRNA is translated. Thus host cell proteins that specifically bind the 5'- and 3'-NCRs of RNA viruses may be potentially required for a process critical for virus replication.

The viral RdRp provides the machinery that will produce complementary RNA copies of the template RNA. It has been shown that relatively pure preparation of several RdRps can replicate several natural or synthetic RNAs. However, RdRp specificity can be brought about by additional viral or cellular proteins indicating an important role of host proteins in RNA virus replication.

A large number and variety of host proteins are increasingly being identified to participate in RNA virus replication/translation. A brief review of these is provided below.

[A] **Host cell proteins interacting with the replicase complex**

The replicase complex of the RNA phage Qβ consists of four subunits: the RdRp and at least three host factors, the protein elongation factor Tu (EF-Tu) and Ts (EF-Ts), and the ribosomal protein S1 (Bluementhal and Carmichael, 1979). Another unknown host
factor, designated as HF, is also required for negative strand synthesis from the positive strand RNA template (Klovins and van Duin 1999; Schuppli et al. 2000). In the generally accepted scenario, the polymerase provides the polymerizing activity, whereas S1 allows the enzyme complex to bind to two internal sites in the viral genome. Removal of these cellular factors resulted in the complete loss of Q8 replicase activity.

Interaction of HCV RNA polymerase (NS5B) with α-actinin was demonstrated by various in vitro techniques and confirmed by reducing the expression of α-actinin by RNA interference in HCV replicon system resulting in decrease of HCV RNA levels (Lan et al. 2003). Similarly, a ubiquitin-like protein hPL1C1 also interacts with NS5B and its over expression in the replicon system reduces the levels of NS5B and of replicon RNA, probably by promoting NS5B degradation via a ubiquitin-dependent pathway (Gao et al. 2003). Using yeast two hybrid system, it was demonstrated that cellular RNA helicase p68 interacts with NS5B of HCV (Goh et al. 2004). It was shown that C-terminal part of NS5B is responsible for this interaction and p68 helicase is redistributed from the nucleus to the cytoplasm in the NS5B-expressing cells.

The RdRp or L protein of Vesicular Stomatitis virus (VSV) interacts with the α-subunit of EF-1 to be partially active, and addition of beta and gamma subunits of EF-1 significantly enhances enzyme activity. All the three subunits are packaged with the L protein in the virions (Das et al. 1998). Tubulin binds to the RdRp (L protein) of Measles virus and may be a part of the polymerase complex of the virus. It is required for efficient RNA replication and transcription since anti-tubulin antibodies inhibit viral RNA synthesis and addition of purified tubulin stimulates MV RNA synthesis in vitro (Moyer et al. 1990).

Eukaryotic initiation factor 3 (eIF3) contains at least 10 subunits and one of its 41 kDa protein subunit has been co-purified with Brome Mosaic virus (BMV) RdRp (Quadt et al. 1993). Similarly, the replication complex of Tobacco Mosaic virus (TMV) contains a 56 kDa protein subunit of eIF3 and the antibodies against this subunit can inhibit TMV RNA synthesis in vitro (Osman and Buck 1997). Hence the host components of two viral polymerases correspond to different subunits of eIF3.
The NSP90 protein, the putative RdRp of Rubella virus, binds to the tumor suppressor retinoblastoma protein, Rb in vitro and in vivo (Atreya et al. 1998; Forng and Atreya 1999). Rb is a nuclear phosphoprotein that binds to many proteins and plays a fundamental role in suppression of various neoplasms. Rb facilitates Rubella virus replication since in null-mutant (Rb-/-) mouse embryonic fibroblasts, the level of Rubella virus replication is lower than in wild-type (Rb+/+) cells (Atreya et al. 1998).

Recently (Uchil and Satchidanandam 2003) while investigating the JEV replication complex using in vitro RdRp assay established evidences of host cellular proteins being part of the RdRp holoenzyme. The identities of these proteins are yet to be established.

[B] Host cell proteins interacting with the viral genomic RNA

The poly (C)-binding protein (PCBP) isoforms 1 and 2 regulate the stability and expression of several mRNAs. They are shown to be involved in Polio virus replication and translation (Gamarnik and Andino 1998). When bound to internal ribosome entry site (IRES) of Polio virus RNA, PCBP enhances translation (Blyn et al. 1997), but it also regulates RNA synthesis since its binding to the disrupted stem-loop IV of the IRES negatively affects replication (Gamarnik and Andino 2000). It also binds to the 5’ cloverleaf structure of the Polio virus RNA upstream of the IRES (Gamarnik and Andino 1997; Gamarnik and Andino 1998; Parsley et al. 1997), to which the Polio virus 3CD polymerase also binds. Using the PCBP-depleted cell extracts, it was shown that PCBP2 rescues both RNA replication as well as translation, whereas PCBP1 rescues only replication (Walter et al. 2002). The ribonucleoprotein (RNP) complex of 5’ cloverleaf structure RNA, 3CD, and PCBP can interact in vitro with the poly (A)-binding protein (PABP), thus circularizing the RNA (Herold and Andino 2001). Such a circularized RNP complex is required to regulate replication and translation of RNA genome. Nucleolin, an abundant nucleoli protein has been shown to re-localize to the cytoplasm and interact very strongly to the Polio virus 3’-NCR (Waggoner and Sarnow 1998).

Using ultraviolet (UV) light induced crosslinking and immunoprecipitation studies, the heterogeneous nuclear ribonucleoprotein (hnRNP) A1 was shown to bind to the terminal repeat sequence (TRS) of the Mouse Hepatitis virus (MHV) RNA of the complementary leader region (Huang and Lai 2001; Li et al. 1997). Moreover, MHV RNA
transcription and replication were stimulated by over expression of hnRNP A1, and inhibited by dominant-negative mutants of hnRNP A1 (Shi et al. 2000). Mitochondrial aconitase binds to the MHV 3'-NCR harboring a poly (A) tail. Binding of this protein is thought to increase the stability of the viral RNA, thereby favoring translation and production of virions at early times after infection (Nanda and Leibowitz 2001).

Poly T binding protein (PTB), a member of the hnRNP family of proteins, shuttles between the nucleus and cytoplasm. It binds to a stretch of four UCUAA repeats in the 5' (+) leader and to the 5'-NCR of the complementary (-) strand of MHV RNA (Huang and Lai 2001). PTB over expression had a dominant negative effect on replication of MHV (Choi et al. 2002). Moreover, PTB interacts in vivo and in vitro with the viral N protein, an element of the replication complex, supporting the idea that PTB is a part of the viral replication machinery (Choi et al. 2002). PTB also interacts specifically with IRES of several picornaviruses enhancing translation. It has been suggested that early in infection, PTB enhances Polio virus translation. However, when Polio virus translation products accumulate, viral protease cleaves PTB and hence during late stages of infection, PTB would no longer be available at concentrations sufficient to favor translation and hence leading to a switch from translation to replication (Back et al. 2002).

A group of three cellular proteins belonging to the “NFAR” group (NF90/NFAR-1, NF45 and RNA helicase A) has been shown to bind to both the 3'- and 5'-NCRs of Bovine Viral Diarrhea virus (BVDV) genome (Isken et al. 2003). Mutation of the proteins' binding sites in the 3'-NCR decreased protein binding and RNA replication. Similarly, transfection of cell lines supporting BVDV replication with siRNA directed against RNA helicase A led to 60-85% reduction of viral replication suggesting that NFAR may mediate circularization of viral RNA leading to a switch from RNA translation to replication.

The 3'-NCR of minus strand of DENV has been shown to interact with calreticulin and PDI. Calreticulin is a ubiquitous calcium storage protein that plays an important role in modulating calcium signals. It also functions as a chaperone in the ER and modulates the expression of hormonally regulated genes. The protein is thought to play a role in DENV replication. The PDI is an abundant protein in the lumen of the ER and catalyses the in vivo folding of a variety of proteins by assisting in isomerization of intramolecular
disulfide bridges (Freedman et al. 1994; Noiva and Lennarz 1992). Both these are important components of the ER, an important compartment for viral translation, replication and encapsidation. Possibly, calreticulin and PDI are involved in the assembly of the replication complex within the 3'-NCR of minus strand of DENV (Yocupicio-Monroy et al. 2003). The phosphorylated form of calreticulin interacts with the SL of the 3'-NCR of the plus-strand of Rubella virus RNA. The RNA binding activity of this form of calreticulin and its affinity for the 3'-end of Rubella virus RNA increases temporally with the appearance of the minus-strand RNA synthesis during infection. These observations suggest that the specific binding of calreticulin to the 3’ plus-SL RNA is important in regulating Rubella virus RNA synthesis (Singh et al. 1994).

WNV is a mosquito-borne virus of the flavivirus genus. The terminal end of the 3'-NCR of genomic RNA of WNV forms a conserved terminal 3'-SL structure. Three proteins of 105, 84 and 56 kDa from the baby hamster kidney (BHK) cells bind specifically to the 3'-SL structure of WNV genomic RNA (Blackwell and Brinton 1995). One of these has been purified and identified as the translation elongation factor EF-1α, which may be involved in targeting WNV RNA into intracellular membranes that provide a microenvironment for the efficient replication of the viral RNA. EF-1α may also interact with viral and other cellular proteins that bind to the WNV positive-strand 3'-NCR, that is, instead of an enzymatic activity, EF-1α may provide protein-RNA and protein-protein interaction that promote the assembly of viral replication complexes (Blackwell and Brinton 1997). The multifunctional TIA and TIAR proteins belonging to a family of RNA binding proteins are known to be involved in translational regulation and apoptosis. These proteins have recently been shown to bind WNV 3’ hairpin in the negative RNA strand (Li et al. 2002). Virus growth is inefficient in TIAR knockout cells and when such cells are complemented with a vector expressing TIAR, WNV growth is partially restored. Thus TIAR and possibly TIA-1 are functionally important for WNV replication. The trailer RNA of Sendai virus also binds TIAR, and here this protein is shown to be involved in virus induced apoptosis (Iseni et al. 2002).

The cellular protein La has been reported to bind to the NCR of positive, negative as well as double-stranded RNA viruses. La protein has been shown to interact with 5' NCR/5'-SL structure of Polio virus (Meerovitch et al. 1993), Rubella virus (Duncan and
La protein is one of the many cellular proteins that bind to the TAR region of the 5'-NCR of HIV RNA (Chang et al. 1994; Svitkin et al. 1994). It is but one of the many cellular proteins found to bind to this RNA sequence; presumably, these proteins together regulate the switching between transcription and translation of HIV RNA. La protein is also known to interact with the 3'-NCR of both positive- and negative-sense genomic RNA as well as 5'-NCR of genomic RNA of DENV and it plays an important role in virus replication. It is believed that the La protein acts as an RNA chaperone to maintain RNA structure and thus protects RNA from rapid degradation. This happens to be a perceivable explanation considering the fact that the virus does not contain a poly-(A) tail. The interaction of La protein with the 3'-NCR (-) might suggest an additional role of La protein to facilitate the interaction of viral RNA with other viral or cellular proteins in order to initiate replication of negative strand. La has an ATP-dependent helicase activity that unwinds double-stranded RNA substrates (Huhn et al. 1997) which is also required for viral replication. La protein interaction with HCV IRES leads to significant level of stimulation of the HCV IRES-dependant translation (Ali et al. 2000; Ali and Siddiqui 1997). Similarly La has been shown to alleviate translation inhibition from the Encephalomyocarditis virus (ECMV) IRES imparted by surplus PTB protein (Kim and Jang 1999). The aberrant translation initiation of Polio virus IRES-mediated RNA is corrected in the presence of La antigen that is also accompanied by a modest stimulation of translation (Meerovitch et al. 1993; Svitkin et al. 1994a). Apart from this, La protein is thought to facilitate the interaction of viral proteins in order to initiate replication of the negative strand. For example in case of Sindbis virus, 44 nucleotides at the 5'-terminus of the genomic RNA can form a stable SL structure. These have been shown to be important for viral replication. The structure formed by the complement of this sequence at the 3'-end of the minus strand RNA has been proposed to be a promoter for RNA replication and was shown to bind cellular proteins of molecular sizes of 42, 44 and 52 kDa by UV
irradiation (Pardigon and Strauss 1992). The 50 kDa protein was demonstrated to be the mosquito homolog of the vertebrate La autoantigen (Pardigon and Strauss 1992).

As for JEV, it has been shown that non-structural proteins NS5 (viral RdRp) and NS3 are involved in RNA replication (Uchil and Satchidanandam 2003). However, it is not clear which host cells proteins, if any, are necessary. Studies in our lab have shown that host proteins bind to RNA representing the JEV 3'-SL structure. At least three proteins of apparent molecular size of 32, 35, 50 kDa from the mouse brain bound to JEV 3'-NCR in UV-induced protein-RNA crosslinking assays. Screening of mouse brain cDNA expression library identified a 36 kDa Mov34 protein that bound to the JEV 3'-SL RNA. This protein belongs to a family of proteins whose members are involved in control of RNA transcription and translation. Mov34 interaction with JEV RNA, thus, may have functional relevance for viral RNA transcription (Ta and Vrati 2000). The identity of the other proteins interacting with JEV 3'-NCR is still not known. In the present work I have shown that La protein binds both the 3'- and 5'-NCRs of JEV with high affinity. This may result in the circularization of genome leading to its efficient transcription and translation and thereby efficient virus replication.

2.5 RNA-Protein Interactions

Organisms utilize a variety of RNA-Protein interactions to regulate the expression of their genes. In fact, most of the functions of RNA involve their interaction with protein/s. RNA binding proteins (RBPs) play key roles in post-translational control of RNAs, which, along with transcriptional regulation is a major way to regulate gene expression. Eukaryotic mRNAs are almost always associated with RBPs that control each and every aspect of RNA metabolism. Understanding how RBPs and RNA interact with each other is, therefore, central to understanding the basal gene expression and its regulation. In the case of RNA viruses, the genomic RNA interacts with several viral and host proteins (see below) and thus understanding this interaction helps us understand the virus replication.

RNA structure and sequence is almost as diverse and complex as those of proteins. This structural diversity defines an enormous variety of shapes and sites for intermolecular recognition. Proteins bind with the affinity required for complex formation at the
concentration of reagents present in living cells and for regulation of biological function. For example, t-RNA synthetase enzyme binds t-RNA with modest affinity ($K_d = \mu M$), while human U1A protein binds to its target RNA much more tightly ($K_d = nM$). In the first case the enzyme would not be effective if the product was not released after catalysis while in second case, it is a permanently assembled structure. Similarly analyses of other RNA-protein interactions reveal that affinity is not simply related to parameters such as charge or size of intramolecular interface area.

RBPs have a modular structure and contain RNA binding domains (RBDs) of 70-95 amino acids that mediate RNA recognition and auxiliary domains that perform additional functions (Biamonti and Riva 1994; Burd and Dreyfuss 1994; Mattaj 1993; Nagai 1996). Based on their amino acid sequence or the structural features the RBDs can be divided into the following groups.

**Arginine-Rich Motif:** In a ‘basic domain’ class of RBPs, RNA recognition is mediated by small, highly basic stretches of 10-15 amino acids rich in arginine and lysine residues (Lazinski et al. 1989). While this sequence is often referred to as a domain and is necessary for RNA recognition in the proteins that contain them, it does not constitute an independent RNA recognition unit in structural or functional sense and hence is not sufficient for RNA recognition. When investigated in isolation, these peptides of 10-15 amino acids bind RNA with comparable affinity to the corresponding protein, but fail to discriminate cognate binding sites from noncognate RNAs (Daly et al. 1995; Weeks et al. 1990). These peptides are unstructured in isolation but their conformation is only defined by the RNA to which they bind (Frankel and Smith 1998). Three most important examples of this group are Tat and Rev proteins of HIV and N protein of bacteriophage λ (Battiste et al. 1996; Legault et al. 1998; Puglisi et al. 1995). How these basic domains function structurally and functionally in the complete context of the protein remains to be elucidated.

**αβ-Protein Domains:** Several proteins contain domains of 60-90 amino acids that are responsible for RNA recognition along with auxiliary domains (Biamonti and Riva 1994; Burd and Dreyfuss 1994; Nagai 1996). These domains are αβ proteins with an
antiparallel β-sheet on one face of the protein packed by a hydrophobic core against an α-helical face. The three most common αβ RNA binding motifs are:

[A] RNA recognition motif (RRM): The RRM is by far the most abundant, diverse and well characterized RNA binding motif (Maris et al. 2005). This domain has a repeated βαβ arrangement (Nagai et al. 1990). Examples of this fold include polyadenylate binding protein (PABP) and heteronuclear ribonucleoprotein (hnRNP) C. Structures of RRM domains in complex with different RNAs show that this small compact domain is a central component of RNA recognition but not the only determinant. N- and C-terminal extensions, multiplication of RRM domains or protein cofactors can play important role in RNA binding specificity. The RRsMs are not only involved in RNA recognition but also in protein-protein recognition.

[B] K-homology (KH) domains: The K-homology domain was originally identified as triple repeat in the hnRNP K protein (Siomi et al. 1993). The domain consists of a stable βαββα fold (Musco et al. 1996). The hydrophobic residues of the three strand cluster on one side of the β-sheet and pack against the hydrophobic amino acids of the three α helices, thus building the hydrophobic core of the structure. The KH-containing proteins play major roles in regulating cellular RNA metabolism. Eukaryotic KH proteins include splicing modulator Mer 1 in yeast (Nandabalan et al. 1993), PSI in drosophila (Siebel et al. 1995), the human RNA associated src-substrate Sam 68 (Fumagalli et al. 1994) and vigilin (Schmidt et al. 1992). There is now considerable evidence that KH domains bind single stranded RNA.

[C] Double stranded RNA binding Domain (dsRBD): As the name implies, dsRBDs have been defined through their ability to bind to double stranded RNA (dsRNA). dsRBDs have an αβββα topology (Bycroft et al. 1995; Kherrat et al. 1995) and is a general dsRNA binding module. Isolated domains bind dsRNA of any sequence with little or no specificity (Clarke and Mathews 1995), but multiple dsRBDs may specifically recognize certain RNA structure (Clarke and Mathews 1995; Ferrandon et al. 1994). These motifs are found in a variety of proteins involved in dsRNA metabolism (Fierro-Monti and Mathews 2000).
Zinc Finger RNA-binding motif: The zinc finger motif was first described as a recurring 30 amino acid repeat that is characterized by four cysteine and histidine residues that surround and chelate a $\text{Zn}^{2+}$ ion. The classical zinc-finger motif (CCHH-type) is one of the most common structural motifs in eukaryotes and accounts for about 3% of genes in the human genome. It is mainly found in DNA-binding proteins such as transcription factors, but some zinc-fingers also bind to RNA and even participate in protein-protein interaction. Compared with the DNA double helix, RNA molecules form complex secondary and tertiary structures including hairpin loops and helical elements closed by hairpin loops, which are recognized by zinc finger domains. This class of zinc finger proteins includes HIV-1 nucleocapsid protein (De Guzman et al. 1998), transcription factor TFIID (Lu et al. 2003) and TIS11d (Hudson et al. 2004).

2.5.1 Differences in RNA and DNA recognition by proteins

Generally DNA-protein recognition occurs by insertion of an $\alpha$-helix into the major groove of the double stranded DNA (Steitz 1990). A specific DNA sequence is then recognized through the formation of extensive hydrogen bonding and Van der Waal’s interactions with the bases (‘direct readout’) and by recognition of sequence-dependent conformational features through electrostatic interactions with the negatively charged phosphodiester backbone of DNA (‘indirect readout’). However, in case of RNA, the major groove of double helical RNA is too narrow to allow insertion of a protein’s $\alpha$-helix or $\beta$-strand. Some DNA binding proteins bind in the minor groove (Kim et al. 1993a; Kim et al. 1993b; Love et al. 1995; Werner et al. 1995) and though the minor groove is well accessible, the chemical groups exposed in the minor groove of nucleic acid bases are not diverse enough among different nucleotides to allow effective discrimination (Seeman et al. 1976). All known sequence-specific RBPs recognize any single-stranded regions and hairpin loops, where the functional groups on the bases become accessible. RNA double-helical regions are recognized only when structural distortions in the double helix generated by internal loops or bulges allow access to the major groove. Further differences exist between DNA-protein and RNA-protein recognition. All DNA double-helical structures are very similar (at least to a first approximation). Thus, sequence-specific DNA recognition requires a very precise reading of the identity of individual nucleotides within the DNA double helix. The diversity of RNA structures favors the recognition of unique
shapes and charge distributions of different RNAs. Thus, despite the chemical similarities between RNA and DNA, many important lessons of DNA-protein recognition are only partially applicable to RNA recognition.

2.5.2 Conformational Flexibility of RNA-Protein Interface and Induced Fit

Hundreds of proteins from higher organisms recognize RNA substrates widely diverse in sequence and structure via RNP domains. Many proteins have multiple copies of RRM, dsRBDs or KH domains. In numerous cases the RNA-binding activity is determined by cooperativity of two or occasionally more domains (Kanaar et al. 1995; Shamoo et al. 1994; Tacke and Manley 1995). A single RNP domain can recognize 5-10 single-stranded RNA bases with high affinity and specificity when the nucleotides are present in a defined RNA structural context. For example, the U1A protein recognizes seven single-stranded nucleotides in the context of a hairpin or internal loop with very high affinity and specificity. When those seven nucleotides are presented in absence of RNA secondary structure, the binding constant is reduced 100000-fold. i.e. almost to the level observed for binding to RNA of random sequence (Hall 1994; Tsai et al. 1991).

The comparison of unbound and bound RNA and protein structures reveals the universal observation of induced fit in RNA-protein recognition. This observation gives rise to the question if conformational changes are mechanistically important for providing specificity but biologically irrelevant feature of RNA recognition or if they are central aspect of biological regulation. Similarly, affinity and hence binding energy is not related to the interface area (presumably a measure of the Van der Waal’s interaction energy and of the entropic contribution to binding energy from solvent and ion release) e.g. as stated above, U1A protein binds tightly to its substrate RNA although the interface area is small while t-RNA synthetase complexes have very large interfaces but bind RNA weakly.

Though numerous high resolution crystal structures of RNA complexes of RBPs are available today, it is very difficult to understand the different specificities observed for distinct RRM proteins. Further knowledge in terms of crystal structure, NMR and computational studies of RNA and protein individually as well as in complex form is needed to understand biological functions of these interactions.
2.6 La protein

La is a conserved, highly abundant, nuclear phosphoprotein that interacts with a large variety of ligands and is the first protein that associates with all newly synthesized RNA polymerase III transcripts via their common UUU-\text{OH} 3' termini. It was first described as an autoantigen (also called SSB) in patients suffering from rheumatic disease systemic lupus erythematosus (SLE) and Sjögren's syndrome (SS) (Alspaugh and Tan 1975; Mattioli and Reichlin 1974). The homologs of La protein have been identified in a variety of eukaryotes (Chambers and Keene 1985; Lin-Marq and Clarkson 1995; Marchetti et al. 2000; Yoo and Wolin 1994). La has been implicated in association with nascent small RNAs like pre-tRNA, pre-5S rRNA, U6 snRNA, RNase P RNA etc. (Chambers et al. 1983; Hashimoto and Steitz 1983; Hendrick et al. 1981; Lerner et al. 1981; Reddy et al. 1983; Rinke and Steitz 1982; Rinke and Steitz 1985; Shen and Maniatis 1982). La is largely confined to the nucleus (Habets et al. 1983; Hendrick et al. 1981; Yoo and Wolin 1994), though cytoplasmic (Habets et al. 1983; Hendrick et al. 1981; Meerovitch et al. 1993) and nucleolar (Deng et al. 1981; Graus et al. 1985; Hendrick et al. 1981) localization has also been demonstrated. In certain circumstances, a fraction of La protein becomes cytoplasmic e.g. during Polio virus infection of mammalian cells, a portion of La protein redistributes to cytoplasm (Meerovitch et al. 1993) and also during apoptosis of mammalian cells (Ayukawa et al. 2000; Casciola-Rosen et al. 1994). Because of change in distribution of mammalian La protein under certain circumstances, the La protein has been proposed to shuttle between nucleus and cytoplasm (Bachmann et al. 1989), however this role is still controversial (Borer et al. 1989).

Alignment of La protein sequences from various organisms reveals that these proteins can be divided into two domains. The highly conserved N-terminal domain (NTD) and the less conserved C-terminal domain (CTD). The NTD comprises of two RRMs. The RRM1 comprises of \(\sim60\) amino acid long "La motif" (Van Horn et al. 1997), a highly conserved sequence that is present in a number of otherwise unrelated proteins. The second RRM, RRM2 is also called as the central RRM and is less well conserved (Bandziulis et al. 1989; Query et al. 1989). The CTD is a highly charged, varying in both size and sequence between species. This portion of the protein contains the third RRM, or C-terminal RRM that is highly basic due to the presence of 40-50% charged residues.
(Aigner et al. 2000; Yoo and Wolin 1994). The human La protein is a multimeric protein of 408 amino acids and contains all three folded domains followed by an unstructured C-terminal region that includes a short basic motif (SBM), a nuclear localization signal (NLS) and a phosphorylation site (Alfano et al. 2004; Jacks et al. 2003; Maraia and Intine 2001; Wolin and Cedervall 2002). The C-terminus also contains a homodimerization domain that is required for the function of La in enhancing the translation of Polio virus (Craig et al. 1997). In lower eukaryotes such as yeast, the protein adopts a simpler form because the RRM3 and SBM are absent (Wolin and Cedervall 2002).

2.6.1 La protein interaction with viral RNAs

Recently La protein has been shown to be engaged by various viruses either by binding their RNAs or by modifying La by proteolysis, to facilitate their own genetic program. La protein has been shown to interact with 5'-NCR/5'-SL structure of Polio virus (Meerovitch et al. 1993), Rubella virus (Duncan and Nakhasi 1997), Influenza virus (Park and Katze 1995), Sindbis virus (Pardigon and Strauss 1996), the HIV-TAR element (Chang et al. 1994), VSV (Kurilla and Keene 1983), Rabies virus (Kurilla et al. 1984), DENV (Garda-Montalvo et al. 2004; Nova-Ocampo et al. 2002; Yocupicio-Monroy et al. 2003), and Coxsackievirus B3 virus (Ray and Das 2002). For the most part, these La interactions with viral RNA elements do not appear to be mediated by the 3'-terminal oligouridylate sequence that are characteristics of La protein prototypical interactions with pre-processed pol III transcripts.

The functional role of La protein binding with viral RNA has been revealed in some of the cases. For example, the aberrant translation initiation of Polio virus IRES-mediated RNA is corrected in presence of La protein that is also accompanied by a modest stimulation of translation (Meerovitch et al. 1993; Svitkin et al. 1994a). La binding to HIV-TAR structure alleviated the translational repression exerted by the TAR on the downstream reporter gene (Svitkin et al. 1994). La protein interaction with HCV IRES leads to significant level of stimulation of the HCV IRES-dependant translation (Ali et al. 2000; Ali and Siddiqui 1997). Similarly, La has been shown to alleviate translation inhibition from the ECMV IRES imparted by surplus PTB protein (Kim and Jang 1999).