CHAPTER 1

Literature Review

1.1 Introduction

Rabies (*rabidus* being Latin for mad) is a highly fatal nervous disease of humans and all other warm-blooded vertebrates. The disease is generally transmitted by the bite of diseased animals, most commonly dogs and other carnivores (Swanepoel *et al.*, 1993). The causative agent of the disease is a member of the rod- or bullet-shaped Rhabdoviridae family (*rhabdos* being Greek for rod) which have a single-stranded, negative sense RNA genomes (Krebs *et al.*, 1995). It is divided into 2 main genera, *Vesiculovirus* and *Lyssavirus*.

Rabies virus is the prototype virus of the genus *Lyssavirus*, and has a large geographic distribution in Africa and Europe (King and Crick, 1988; Bourhy *et al.*, 1993). Control of the disease and epidemiological monitoring of the virus have historically been the main focus of rabies virus research. The epidemiology of the disease in South Africa has been studied in some detail over the last decade. Rabies was found to be epizootic in domestic dogs in KwaZulu-Natal, with little wildlife involvement. This is in contrast to the rest of the country where three wildlife host species, *viz.* the yellow mongoose, bat-eared fox and black backed jackal (each associated with a specific geographical region) act as reservoir species for rabies virus (von Teichman *et al.*, 1995). Based on serotypical and genotypical data, South African rabies viruses exist in two groups: the canid viruses that infect domestic dogs, jackals and bat-eared foxes and the viverrid virus group consisting of viruses circulating in viverrid hosts (mongooses and genets).

The genus *Vesiculovirus* is essentially comprised of the prototype vesicular stomatitis virus (VSV). The virus has served as a model for much of the research regarding members of the Rhabdoviridae family. This has been due to the high degree of similarity in genomic organization, as well as transcription and replication cycles between the members of this virus family. For VSV, the viral genes have been cloned and the regulation of gene expression has been extensively studied *in vitro*. These studies have been facilitated by the development of infective cDNA clones, which enabled the controlled manipulation of gene expression.

Despite their similar genomic organization (Abraham and Banerjee, 1976; Flamand and Delagneau, 1978), as well as patterns of RNA and protein synthesis during the infectious cycle

(Coslett *et al.*, 1980; Holloway and Obijeski, 1980), rabies virus and VSV exhibit notable differences in their biological activity.

- Rabies virus requires nuclear functions (Wiktor and Koprowski, 1974), but VSV is capable of infecting enucleated cells (Follet *et al.*, 1974).
- The rabies virus infectious cycle is very slow with viral protein synthesis not observed until 6 h post-infection (Hummeler *et al.*, 1967). In contrast, VSV infection is rapid and highly efficient with progeny virions emerging as early as 4 h post-infection (Simonsen *et al.*, 1979).
- Unlike rabies virus that continues to produce progeny virions for several days post-infection (Hummeler *et al.*, 1967; Matsumoto, 1974), VSV rapidly inhibits host cell macromolecular synthesis, (McGowan and Wagner, 1981; Weck and Wagner, 1978; Wertz and Youngner, 1972) and most of the infected cells are killed at 12 h post-infection (Marcus and Sekellick, 1975; Marcus *et al.*, 1977; Weck and Wagner, 1979).

1.2 Structure of Rhabdoviruses

Rabies virus has a bullet-shaped morphology (Fig.1A) and the virions consist of a nucleocapsid core surrounded by a host-derived lipid envelope. The nucleocapsid core contains all the elements necessary for viral transcription (Kawai, 1977). These include the single-stranded, negative sense RNA genome that is associated with the nucleoprotein (N), the large polymerase protein (L) and the phosphoprotein (P) (Kawai, 1977). In addition to these proteins, two membrane proteins, the matrix protein (M) and glycoprotein (G), are located within the virus envelope (Fig. 1.1B).

1.2.1 The viral genome

Rabies viruses contains a single-stranded, negative sense RNA genome which is 11 932 nucleotides in length, in the case of the Pasteur virus (Tordo and Poch, 1988). Transcription of the rabies virus genomic RNA, as with VSV, produces sequentially one leader RNA (Kurilla *et al.*, 1984) and five polyadenylated monocistronic mRNAs (Coslett *et al.*, 1980; Holloway and Obijeski, 1980).

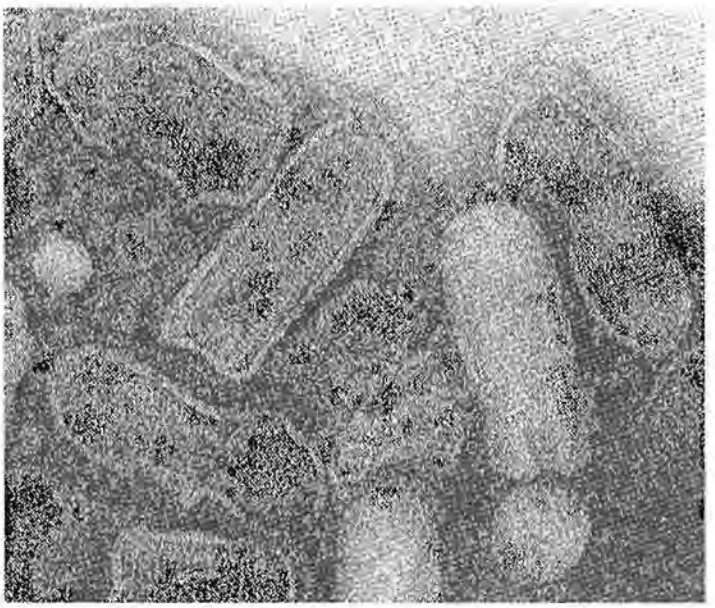


Fig. 1A. Electron micrograph of rhabdovirus particles (VSV) indicating the characteristic bullet-shaped virus particles (Fraenkel-Conrat *et al.*, 1988).

The gene order along the genome, from 3' to 5', is as follows: leader, N, P, M, G and L (Flamand and Delagneau, 1978) (Fig. 2). A conserved intergenic region consisting of the dinucleotide, GA, separates the five protein-encoding genes in the case of VSV (Rose, 1980). In contrast, the four intergenic regions found in the rabies virus genome are of variable length and composition. There are 2 nt separating the nucleoprotein gene from the phosphoprotein gene, 5 nt each separating the phosphoprotein gene from the matrix protein gene and the matrix protein gene and the glycoprotein gene, and 24 to 29 nt separating the glycoprotein gene from the large polymerase gene (Conzelmann *et al.*, 1990, Tordo *et al.*, 1986). These intergenic regions play an important role during transcriptional attenuation at the gene borders, resulting in the mRNA being produced in a sequential gradient (Finke *et al.*, 2000).

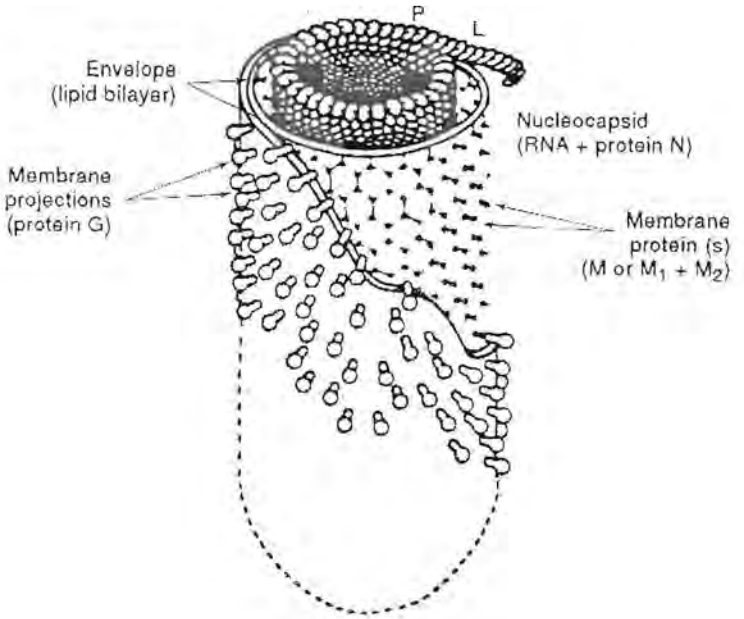


Fig. 1.1B. Schematic representation of the rabies virus virion showing the helical nucleocapsid surrounded by a host-derived lipid bilayer containing the matrix protein and the transmembrane glycoprotein (Wunner, 1991).

The 3' terminus of the VSV genome is 50 to 53 nt in length (Giorgi *et al.*, 1983) and plays a critical role in regulatory steps such as polymerase initiation (Emerson, 1982; Emerson and Yu, 1975; Keene *et al.*, 1978) and N protein encapsidation (Blumber *et al.*, 1983). It also acts as a switch from transcription to replication (Blumberg *et al.*, 1981). The rabies virus leader RNA is 56 nt in length (Kurilla *et al.*, 1984). Blumberg *et al.* (1983) have proposed that the first 14 nucleotides in the VSV leader RNA segment, which contains an A residue on the first and every third base (ACG AAN ACN ANN AAA), constitutes the encapsidation initiation signal. In rabies virus, the encapsidation initiation sequence has been identified as the A-rich sequence spanning nucleotides 20 to 30 in the leader RNA (5'-AAGAAAAAACA-3') (Yang *et al.*, 1998). A similar encapsidation initiation sequence (5'-AAAAATGAGA-3') is present at nucleotides 20 to 30 at the 5' end of the negative-strand genomic RNA (Conzelman *et al.*, 1990; Yang *et al.*, 1998).

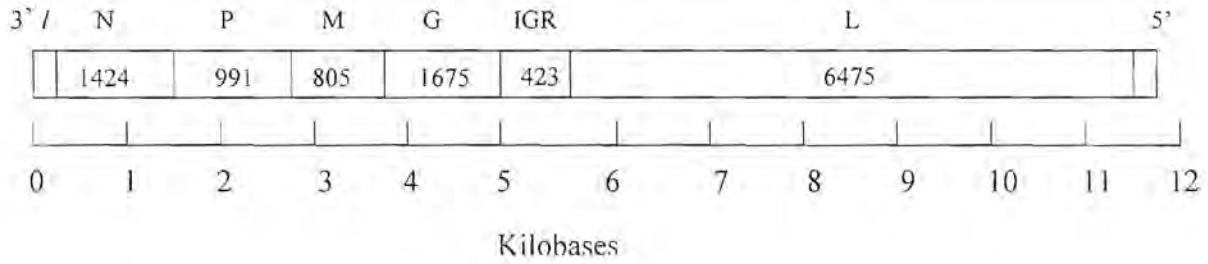


Fig. 1.2. A schematic representation of the single-stranded negative sense RNA genome of rabies virus. The order of each genomic region, in a 3' to 5' direction is: *l* – leader RNA; N – nucleoprotein gene; P – phosphoprotein gene; M – matrix gene; G – glycoprotein gene; IGR – intergenic region; L – large polymerase gene. The length of the respective genes (nucleotides) is indicated in each gene.

1.2.2 Rabies virus proteins

1.2.2.1 Glycoprotein (G)

Primarily as a result of its importance in vaccine-development, the glycoprotein has been extensively studied. The glycoprotein is composed of four distinct domains: the signal peptide that allows the translocation of the polypeptide through the membrane of the endoplasmic reticulum (Tordo and Kouknetzoff, 1993), the ectodomain that is exposed on the outer surface of the virion and includes the glycosylation, palmytolation and antigenic sites (Coulon *et al.*, 1993), the transmembrane peptide that anchors the protein within the viral envelope (Tordo and Kouknetzoff, 1993), and the cytoplasmic domain located in the inner part of the virion (Tordo and Kouknetzoff, 1993). Both the signal peptide and transmembrane peptide are hydrophobic, while the cytoplasmic domain is hydrophilic. The transmembrane glycoprotein stimulates the humoral and cell-mediated immunity against viral infection by not only inducing and binding the virus neutralizing antibodies, but also by stimulation of the T cells (Tordo and Poch, 1988). Furthermore, the G protein plays an important role in the virus-host cell interaction in that it mediates attachment of the virus to the host cells (Tordo and Poch, 1988).

1.2.2.2 Matrix protein (M)

The matrix protein is the most abundant protein present in the virions. The protein plays a central role in viral assembly and release of the viral progeny from infected cells by budding. Cells infected with a rabies virus deficient in M protein were shown to lack the typical bullet-shaped morphology of rhabdovirus particles. Instead it contained long, rod-shaped virions, thus

demonstrating severe impairment of the virus assembly process. Complementation of the M protein-deficient virus by recombinant plasmid rescued virus formation (Mebatsion *et al.*, 1999). In infected cells, 10% of the M protein population is stably associated with the plasma membrane, whereas most of the cellular M protein is found in soluble form in the cytoplasm up to the moment of incorporation into new virions (Barge *et al.*, 1996). The interaction between the G protein and the RNP complex is mediated by a layer of M protein molecules, allowing efficient release of infectious progeny virions (Mebatsion *et al.*, 1999).

1.2.2.3 RNA polymerase protein (L)

The L protein gene is 6 475 nucleotides long (HEP-Flury strain) and encodes for a polypeptide of 2 127 amino acids, with a molecular weight of 242 938 Da (Morimoto *et al.*, 1998). The L protein is a multifunctional enzyme and is the RNA-dependent RNA polymerase. This protein reportedly carries out all of the enzymatic steps required for transcription, including initiation and elongation of transcripts as well as co-transcriptional modifications of RNAs such as capping, methylation and polyadenylation (Banerjee, 1987).

1.2.2.4 Nucleoprotein (N)

The N protein of rabies virus is responsible for the encapsidation of the RNA genome, ensuring the genome's protection against nucleases. The N protein binds to the phosphate-sugar backbone of the RNA and exposes the nucleotide bases to the RNA polymerase during transcription and replication (Emerson, 1987; Iseni *et al.*, 2000; Keene *et al.*, 1981; Klumpp *et al.*, 1997). In addition, the N proteins of all Rhabdoviruses are thought to play a crucial role in transition of the RNA synthetic mode from transcription to replication (Patton *et al.*, 1984; Wertz *et al.*, 1987).

The N protein of rabies virus has been expressed in recombinant baculovirus-infected cells, purified and characterized in great detail (Iseni *et al.*, 1998). The baculovirus-expressed N protein had the same morphological characteristics as viral nucleocapsids when observed with negative stain electron microscopy (EM). The nucleocapsids were 84 Å long; 53 Å high and the spacing between the N protein monomers along the coil were 35 Å. The recombinant nucleocapsids contained cellular RNA and treatment with RNase yielded a 9 nucleotides per N

monomer stoichiometry, similar to that found in viral nucleocapsids. The same group of researchers also studied the rabies virus N-RNA rings by cryo-electron microscopy and calculated three-dimensional (3D) models from single-particle image analysis combined with back projection. It was found that the N protein has a bi-lobed shape, and each monomer has two sites of interaction with each neighbour [(Fig. 1.3) (Schoen *et al.*, 2001)]. Trypsin treatment of the N protein cuts at lysine-376 and removes a 17 kDa fragment at the C-terminal of the N protein (Iseni *et al.*, 1998; Kouznetzoff *et al.*, 1998). The 3-D structure after trypsin digestion indicated that it cuts off part of one of the lobes without shortening the protein or changing other structural parameters (Fig. 1.3). However, this digested protein was unable to bind to the P protein, indicating that the C-terminal part of the N protein is the only binding site for the P protein on the N-RNA.

RNA-binding studies have revealed that the N protein binds 5 to 10 times more leader RNA than to non-leader RNA (Yang *et al.*, 1998). Furthermore, N protein encapsidation of non-leader RNA, but not the leader RNA, was inhibited when the P protein was simultaneously added to the encapsidation reaction, indicating that the P protein helps confer the specificity of leader RNA encapsidation by the N protein (Yang *et al.*, 1998). A hyperconserved motif within rabies virus N proteins starting with the sequence NH_2 S-P-Y-S-S-N (position 298-352) has been identified as being primarily responsible for binding of the leader RNA (Kouznetzoff *et al.*, 1998). However, the conformation of the N protein may also be important in allowing N protein-RNA interaction (Kouznetzoff *et al.*, 1998). This has been supported by the following observations: temperature has an inhibitory effect on the RNA-binding specificity of the N protein, N protein solubilized from purified RNPs or in bacterial extracts following treatment with guaniduim HCl, resulted in a dramatic decrease in RNA-binding activity, and limited structural changes in the NH_2-terminal core had a negative effect on its RNA-binding activity.

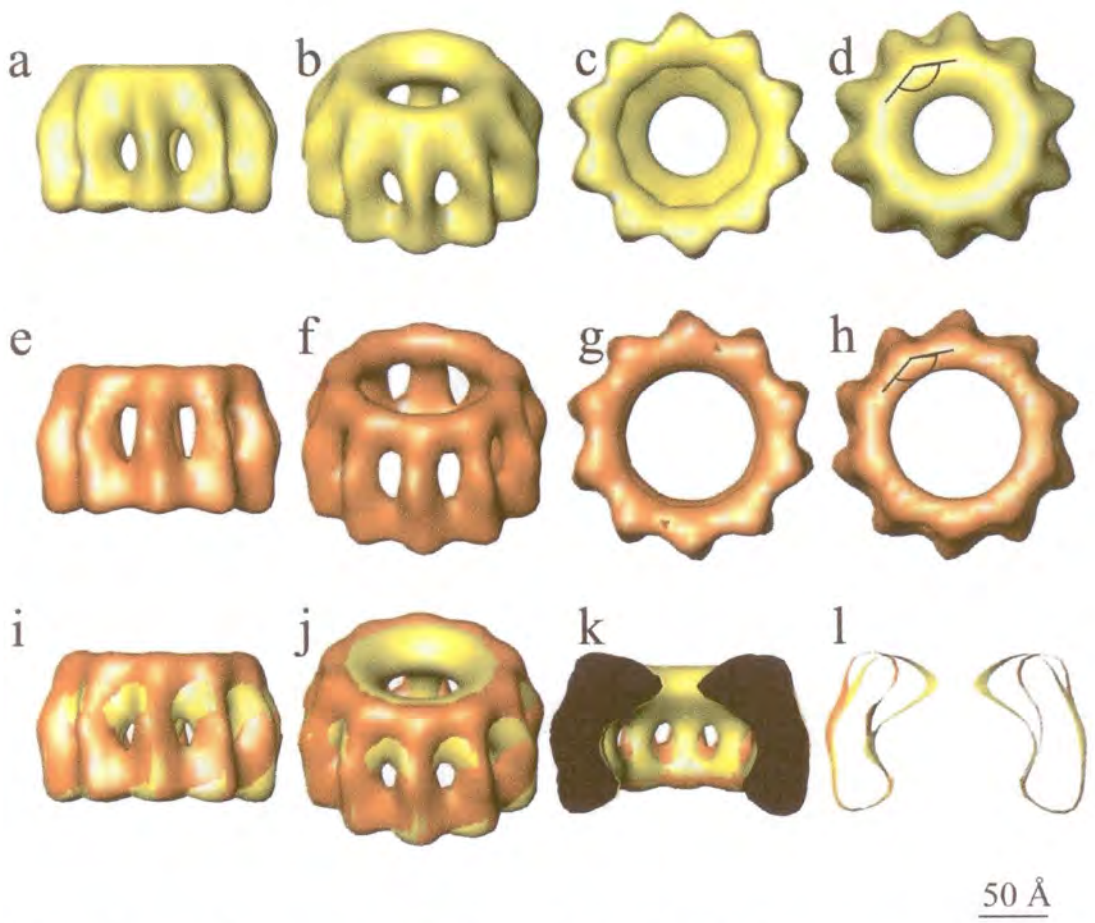


Fig. 1.3. Three-dimensional reconstruction of native (yellow) and trypsin digested (orange) N-RNA rings. The maps are shown in side views (a and e), tilted views (b and f), and end views from the bottom (c and g) and form the top (d and h). The lower panel shows a superposition of the native and digested rings in side view (i), tilted view (j), with the front half cut off (k), and in axial slice (l). The two structures are very similar except at the top end, where the inner part of the circle is missing for the digested rings. The angle between N-N-N shown in panels d and h measures 144° (reproduced from Schoehn *et al.*, 2001).

1.2.2.5. Phosphoprotein (P)

Although the rabies virus P protein is abundant in virus infected cells, it presents only 6% of the virion (Kawai, 1977). Chenik *et al* (1995) reported that the P protein exhibits a large degree of heterogeneity. The authors detected the P protein and at least four additional shorter products (P2, P3, P4 and P5) in infected BSR cells, in BHK cells transfected with a plasmid encoding the wild-type P protein and in purified virions of the CVS rabies virus strain. Mutagenesis of the 5'-terminal region of the P mRNA initiation codon indicated that these shorter proteins were the result of ribosomal leaky scanning (Chenik *et al.*, 1995). In the proposed translation model, ribosomes bind at the capped 5' end of the mRNA and scan downstream to the first AUG codon. Failure to initiate translation at this codon would result in continued scanning downstream for the next available AUG; a second failure allows recognition at a third AUG (Kozak, 1986). Immunofluorescence staining experiments performed on transfected cells have indicated that three of these shorter products (P3, P4 and P5) initiated from, respectively, the third, fourth, and fifth AUG codons. P3, P4 and P5 were mostly located within the nucleus, whereas the wild-type P protein and P2-form were mainly cytoplasmic. Thus, a truncation of at least the N-terminal 52 amino acids of P protein appears to result in the ability to enter the nucleus. A possible explanation is that the N-terminal deletions remove a cytoplasmic retention signal. It is unclear what the function of these truncated P proteins may be, but the presence of large amounts of P2 and P3 forms in purified virions suggests that they are involved in one of the steps of viral multiplication in infected cells (Chenik *et al.*, 1998).

Within non-segmented negative sense RNA viruses, nucleotide sequence analysis of the P protein reveals very little amino acid sequence similarity (Banerjee and Barik, 1992; Lamb and Kolakofsky, 1996). The protein also varies in size, ranging from 27 to 68 kDa between different viruses. Despite these differences, these proteins are all highly acidic and heavily phosphorylated. During viral gene expression the P protein plays a dual role by acting as an essential component of the polymerase (P-L complex), and as a chaperone to deliver soluble N protein (P-N complex) for the encapsidation of RNA during replication (De and Banerjee, 1997). Three functional domains have been identified in the P protein of *Rhabdoviridae*, using VSV as the prototype. Mutational and deletion analyses of the P protein of VSV have indicated at least three functional domains, which play distinct roles during transcription. Domain I is a

highly acidic region that is located at the amino-terminus of the protein (Gill and Banerjee, 1985; Hudson *et al.*, 1986) and must be phosphorylated by cellular casein kinase II (CKII) (Barik and Banerjee, 1992a) at specific sites to serve as a transcriptional activator (Chattopadhyay and Banerjee, 1988; Takacs *et al.*, 1992; Barik and Banerjee, 1992b; Gao and Lenard, 1995a; Chen *et al.*, 1997). VSV specifically packages CKII that remains strongly associated with the ribonucleoprotein complex during morphogenesis (Gupta *et al.*, 1995). Mutational analyses of the phosphorylated amino acids suggested that phosphorylation of this domain may play a critical role during transcription (Pattnaik *et al.*, 1997), but it is not essential for replication (Spadafora *et al.*, 1996; Pattnaik *et al.*, 1997). This view is further supported by the findings that phosphorylation-negative P mutants efficiently form soluble N-P complexes (Takacs *et al.*, 1993; Takacs and Banerjee, 1995) and can initiate the replication process *in vitro* in the presence of cellular factors (Gupta and Banerjee, 1997). This domain also interacts with the L protein to form the active RNA polymerase (Emerson and Schubert, 1987; Takacs and Banerjee, 1995). Domain II is phosphorylated by an L protein-associated kinase (LAK) (Chattopadhyay and Banerjee, 1997b) and is also involved in the binding of P protein to the L protein (Gill *et al.*, 1986; Paul *et al.*, 1988). Mutational analyses of this domain have indicated that phosphorylation is essential for transcriptional activity *in vitro* (Chattopadhyay and Banerjee, 1987b; Gao and Lenard, 1995b). Domain III is a basic and highly conserved 21-amino acid region at the C-terminal of the P protein. This domain is required for binding to the N-RNA template (Gill and Banerjee, 1985; Gill *et al.*, 1986; Emerson and Schubert, 1987) as well as for interaction with soluble N protein (Takacs *et al.*, 1993; Takacs and Banerjee, 1995), indicating its important role in both transcription and replication.

1.3 The processes of viral transcription and replication

From the preceding descriptions, it is apparent that the rabies virus N and P proteins play an important role in the viral transcription and replication processes. Therefore, these processes will be reviewed with reference to the role of the N and P proteins in these processes. Since the discovery of Baltimore *et al.* (1970) that VSV packages the RNA-dependent RNA polymerase that transcribes the genome RNA into messenger RNAs within the mature virion, VSV has became the prototype model for non-segmented negative sense RNA viruses such as members of the Rhabdoviridae and Paramyxoviridae families. While the gene order and overall

mechanism of gene expression in these two virus families are very similar, specific differences have been found and will be discussed in the following sections.

1.3.1 Viral Transcription

The VSV genome is tightly complexed with the nucleoprotein (N) to form the nucleocapsid structure. The active template for transcription and replication of the VSV genome is the nucleocapsid structure, which together with the large polymerase protein (L), and the phosphoprotein (P) serves as the active template for transcription and replication. Both transcription and replication is catalyzed by the viral RNA-dependent RNA polymerase complex, which is composed of the L and P proteins (De and Banerjee, 1985; Emerson and Yu, 1975; Naito and Ishihama, 1976). Biochemical and genetic studies suggest that the L protein is responsible for the enzymatic activities, including the polymerase activity (Hercyk *et al.*, 1988; Hunt *et al.*, 1984; Schubert *et al.*, 1985). Although no enzymatic activities have been associated with the P protein, the phosphorylation status of this protein has been shown to influence the transcriptase and replicase functions of the polymerase protein (Emerson and Schubert, 1987; Pattnaik *et al.*, 1997).

The promoter for VSV genome transcription consists of three separate elements, the 3' leader region, the non-transcribed leader RNA-N gene junction, and the N gene start signal (Whelan and Wertz, 1999a). Additional distinct sequences within the leader region (nucleotides 19 to 29 and 34 to 46) are required for transcription but not replication (Whelan and Wertz, 1999a; Li and Pattnaik, 1999). The RNA polymerase initiates transcription from the 3' terminal end of the genome (Emerson, 1987) and transcribes the genes in a sequential manner (Abraham and Banerjee, 1976; Ball and White, 1976). In doing so, the viral RNA polymerase first synthesizes a small 47-nucleotide long leader RNA that is uncapped and non-polyadenylated and does not encode any viral protein. Following the leader RNA synthesis, five capped and polyadenylated mRNAs, encoding each of the individual viral proteins, are transcribed sequentially in the same order that these genes are found in the virus genome (3'-N-P-M-G-L-5') (Abraham and Banerjee, 1976; Ball and White, 1976).

Schnell *et al.* (1996) have demonstrated that a sequence of 23 conserved nucleotides, which is found at the 3' and 5' junction of all VSV genes, could direct the expression of foreign genes within recombinant VSV genomes. This provided the first evidence that these sequences contained all the *cis*-acting signals necessary to polyadenylate and terminate transcripts from the upstream gene and then to reinitiate transcription at the adjacent downstream gene. The conserved sequence, 3'-AUAC$(U)_7$-5', which is found at the 3' ends of each gene is critical for both polyadenylation and termination of VSV transcription (Barr *et al.*, 1997a, Hwang *et al.*, 1998). When this sequence is encountered, the polymerase stutters over the seven U nucleotides to produce a poly(A) tail approximately 150 nucleotides in length (Schubert *et al.*, 1980). Immediately following the polyadenylation signal, there are two non-transcribed intergenic dinucleotides that may be required for transcription termination (Stillman and Whitt, 1997). However, other studies have suggested that it may function as a spacer element between the transcription termination and re-initiation signals (Barr *et al.*, 1997b; Hwang *et al.*, 1998). Following the intergenic dinucleotide is the sequence 3'-UUGUCnnUAC-5' (with n being any nucleotide). The first three nucleotides of this sequence are the most important for efficient gene expression, as mutations at these positions severely reduce the amount of mRNA levels from the mutated gene. This sequence may be important for re-initiation following polyadenylation and release of the upstream mRNA (Stillman and Whitt, 1997).

There exists a gradient in the molar amounts of the transcribed mRNAs that also follows the gene order from the 3' end of the genome so that the 3'-proximal N gene is transcribed most frequently and the 5'-distal L gene is transcribed least frequently. The gradient in the molar amounts of the mRNAs is believed to be due to attenuation at each of the gene junctions during transcription (Iverson and Rose, 1981). This may be the result of the inability of the polymerase to reinitiate transcription of the downstream gene following transcription termination and polyadenylation of the upstream mRNA.

1.3.2 Replication

1.3.2.1 Switch between transcription and replication

During replication, the transcriptase enzyme switches to the replicative mode and transcribes across the gene boundaries, synthesizing the full-length complement of the negative-strand

genome RNA. The P-L protein complex responsible for transcription is modified to a replicase complex when it reaches the leader-N gene junction (Vidal and Kolakofsky, 1989; Kolakofsky *et al.*, 1991). The modification is presumably triggered by binding of newly synthesized N protein, likely in the form of a soluble P-N protein assembly complex, to nascent leader RNA (Leppert *et al.*, 1979; Blumberg *et al.*, 1981; Peluso and Moyer, 1988). Since N and P genes are the most proximal to the 3' end of the transcribing genome, synthesis of N and P proteins results in immediate complex formation, thus keeping N protein in a soluble and replication-competent form (Bell *et al.*, 1984). The P protein, once complexed with the N protein, is not required to be phosphorylated in Domain I by casein kinase II for its replication function, whereas phosphorylation of the P protein is necessary to interact with the L protein to form the active RNA polymerase complex (Gupta and Banerjee, 1997). As the concentration of N-P protein complex rises in the cell following infection, it acts on the RNP complex to down-regulate transcription and in association with a putative host protein(s) switches transcription to replication (Gupta and Banerjee, 1997). This model is appealing in that all available polymerase activity is channeled toward synthesis of mRNAs until a sufficient amount of the capsid proteins accumulates for replication. The N protein in the complex enwraps the nascent RNA chains of both positive and negative sense genome-length RNAs. As the concentration of N-P protein complex falls below a threshold value, replication ceases and transcription ensues, and this cycle continues during the life cycle of the virus.

1.3.2.2 Replication

During replication, the RNA polymerase first synthesizes the full-length positive sense antigenome, which is enwrapped with the N protein, forming the N-RNA complex. This complex then serves as the template for the synthesis of the negative sense progeny genome RNA (Banerjee, 1987; Banerjee *et al.*, 1977). It is envisaged that the N protein complexes with the nascent leader RNA transcript to initiate encapsidation of the growing RNA chains, leading to replication (Banerjee, 1987; Blumberg *et al.*, 1983; Blumberg *et al.*, 1984; Blumberg *et al.*, 1981; Giorgi *et al.*, 1983).

It is generally believed that the 3'-terminal RNA sequence of the genome RNA is the binding site of the VSV RNA polymerase (Banerjee *et al.*, 1977; Isaac and Keene, 1982; Keene *et al.*,

1981) to initiate transcription. Thus, the 3'-terminal domain of the genome RNA and its complement (leader-sense) RNA are the two important *cis*-acting RNA sequences that are potential targets for cellular proteins to bind and promote the transcriptase and the replicase functions, respectively. It has been shown previously that both positive sense and negative sense leader RNA (the complement of the 3'-terminal sequence of the positive sense genome RNA) interact specifically with the nuclear autoantigen, La protein, in infected cell cytoplasm, raising the possibility that this interaction may have some specific role in the replicative pathway of the virus (Kurilla and Keene, 1983; Kurilla *et al.*, 1982; Wilusz *et al.*, 1983). Moreover, in view of the similarity in the sequences of RNA polymerase III products and the 3' end of the leader RNA, it seems that the interaction of La protein with the leader RNA may be mediated by a sequence motif which regulates VSV transcription and replication. In a separate series of studies, the leader RNA of VSV was implicated in inhibiting cellular RNA synthesis by its temporary localization inside the nucleus following infection (Kurilla *et al.*, 1982; Weck and Wagner, 1978), suggesting that it may interact with specific nuclear proteins involved in RNA synthesis. McGowan *et al.* (1982) tested this by using a soluble cell extract as the source of the RNA polymerase, and reported that purified leader RNA inhibits DNA-dependent transcription of adenovirus and simian virus 40 genes *in vitro*. Gupta *et al.* (1998) subsequently indicated binding of the leader RNA of VSV with hnRNP U and its localization in the cytoplasm. As a result, they speculated that the leader RNA, by its binding to hnRNP U, may be involved in VSV-mediated shut-off of host DNA and RNA metabolism, *via* structural alteration of the leader-N gene junction, enabling the RNA polymerase to read through the junction region and leading to the synthesis of full-length antigenome.

1.4 Aims of this study

From the literature review, it is clear that the nucleoprotein of negative-sense, single stranded RNA viruses is responsible for encapsidating the viral genome, consequently controlling host and viral protein access to the virus genome, and effectively regulating virus transcription and replication processes. Rabies virus differ from the prototype member of the *Rhabdoviridae* virus family, Vesicular stomatitis virus (VSV), in that the nucleoprotein (N) and the phosphoprotein (P) are both phosphorylated, in contrast to the P protein being the only phosphorylated protein of VSV. Despite the insoluble nature of the N protein, the

characterization and functional significance of phosphorylation of the rabies virus N and P proteins were the focus of many published reports in recent years.

The aims of this study can be summarized as follows:

1. Cloning and characterization of the N and P proteins of a South African rabies virus isolate.
2. Expressing both the N and P proteins, singularly and in combination in a BAC-to-BAC™ baculovirus expression system.
3. Expressing the N and the P proteins in *Escherichia coli* using a pGEX expression system.
4. Developing an *in vitro* phosphorylation strategy.
5. Investigating the influence of phosphorylation of the N protein on its ssRNA binding ability, as well as characterization of the specificity of N protein binding to ssRNA.
6. Characterize the influence of phosphorylation on complex formation between the N and P proteins.

CHAPTER 2

Genetic characterization of the nucleoprotein and phosphoprotein genes and encoded gene products of a South African viverrid rabies virus strain

2.1 Introduction

Many studies of viruses in the *Lyssavirus* genus, of which rabies virus is the prototype virus, have focused on the nucleo- and glycoproteins. This has been due to their high antigenicity, which has led to the production of monoclonal antibodies for use in of rabies virus strain differentiation. The phosphoprotein has traditionally received less attention. However, this situation has changed when the important role of the phosphoprotein, in combination with the nucleoprotein, during the process of viral transcription and replication was realized.

The rabies virus nucleoprotein is encoded by the N gene and is 1 424 nucleotides in length (Pasteur virus) (Tordo *et al.*, 1986b). The nucleoprotein plays an important role during virus transcription and replication. Not only is it capable of binding to the viral RNA genome thereby protecting it from nuclease digestion (Kouznetzoff *et al.*, 1998), but it also forms complexes with the phosphoprotein (Schoenh *et al.*, 2001), which plays an important regulatory role during virus propagation. Comparative sequence analyses between different rabies virus strains indicated high levels of homology, e.g. 97.5% between the N genes of Pasteur virus (PV) and CVS; and 99.1% between the N genes of PV and SAD-B19 (Tordo *et al.*, 1986a; Conzelmann *et al.*, 1990). This high level of sequence conservation may point to the importance of maintaining the nucleocapsid structure, which is believed to have been preserved during evolution of the Rhabdoviruses (Bourhy *et al.*, 1989).

The phosphoprotein-encoding gene is 900 nucleotides in length and is highly conserved between different rabies virus strains, displaying 99% homology between the P genes of the Pasteur virus, ERA and PM strains (Larson and Wunner, 1990). Differences between P protein amino acid sequences tend to cluster to specific regions (Nadin-Davis *et al.*, 1997). The N-terminal 50 amino acids is well conserved in all rabies virus strains, the next 35 amino acids (residues 51 to 85) is more variable, while a second variable region is located at position 130 to 180. The P protein is complexed with the N protein as well as the large polymerase protein and thus forms part of the nucleocapsid core (De and Banerjee, 1997). The P protein has been shown to bind to the L protein (Chenik *et al.*,

1998) thereby forming the active RNA polymerase in which the P protein serves as a non-catalytic co-factor (Emerson and Schubert, 1987).

Both the rabies virus N and P proteins are phosphorylated. Phosphorylation of the rabies virus N protein is one of the features that distinguishes it from the cognate VSV N protein. In contrast to the N protein that is phosphorylated at serine and threonine amino acid residues (Anzai *et al.*, 1997; Yang *et al.*, 1999), the P protein is phosphorylated only at serine amino acid residues (Gupta *et al.*, 2000). Although the kinases responsible for rabies virus P protein phosphorylation has been identified as protein kinase C (PKC) and rabies virus protein kinase (RVPK) (Gupta *et al.*, 2000), no information regarding the kinase(s) responsible for phosphorylation of the N protein is available. Phosphorylation of the N protein has been reported to result in less leader RNA being encapsidated and in up-regulation of viral transcription and replication processes (Yang et al., 1999). Phosphorylation of the P protein may play a role in determining the specific host range of the rabies virus (Prehaud *et al.*, 1992; Gupta *et al.*, 2000).

The objectives of this part of the study were to clone and sequence the full-length nucleoprotein and phosphoprotein genes of a South African viverrid rabies virus, and to compare the obtained sequences to the homologous genes and encoded proteins of other rabies virus strains in order to identify conserved features which may be related to the function of the respective proteins.

2.2 Materials and Methods

2.2.1 Monoclonal antibody typing of Rabies virus isolates

Viverrid rabies isolate m710/90 was obtained from the Rabies Unit of the Veterinary Institute at Onderstepoort. This virus was isolated from a rabid *Cynictis Penicillata* (Yellow mongoose) in 1990 and characterized as follows: monoclonal antibody (Mab) typing was performed using a Mab-N panel consisting of 2 Mab-Ns from the Wistar Institute, Philadelphia, and 4 Mab-Ns prepared at the Central Veterinary Laboratory (CVL), Weybridge. Wistar Institute Mab-Ns were used at the dilutions recommended by the donors and the Weybridge Mab-Ns were used at a concentration of four-fold their fluorescence end-point, as previously described (King *et al.*, 1993). Sequence analysis has also found this rabies virus isolate (m710/90) to belong to the viverrid biotype (Olivier, 1997). A 10% suspension of original brain material was inoculated into 21-day-old suckling mice and following virus multiplication, the rabies virus isolate was stored in a 20% lyophilized mouse brain suspension.

2.2.2 Design of nucleotide primers for amplification of the Nucleoprotein (N) and the Phosphoprotein (P) genes

To amplify the full-length N and P genes, oligonucleotide primers were designed based on nucleotide sequence alignments of the 5' and 3' terminal regions of the corresponding genes from various rabies virus strains. The ClustalX multiple sequence alignment program was used for this analysis (Thompson *et al.*, 1997). In order to facilitate cloning of the PCR-amplified genes, unique restriction endonuclease recognition sites were also included in the primers (Table 2.1).

2.2.3 Total RNA extraction

Total RNA was extracted using TRI REAGENT® (Molecular Research Center, Inc.) according to the manufacturer's instructions. Briefly, *ca.* 50 – 100 mg of lyophilized mouse brain material was homogenized in 1 ml of TRI REAGENT®. Following addition of 200 µl of chloroform, the homogenate was separated into aqueous and organic phases by centrifugation at 12 000 rpm for 15 min at 4°C. The RNA was precipitated from the aqueous phase by addition of 500 µl of isopropanol. Following centrifugation, as above,

the pellet was washed twice with 75% ethanol and then suspended in diethyl pyrocarbonate (DEPC) H_2O.

Table 2.1 Primers used to amplify and sequence the nucleoprotein (N) and phosphoprotein (P) genes.

Gene	Primer	Use	Sequence (5' – 3')	Reference
N gene	JJ1	cDNA, PCR, sequencing	ggatcctacaatggatgccgac	This study
	N_3	PCR, sequencing	**gaattc**gatgtctggcgtcttgcc	This study
	N_2	Sequencing	cccatatagcatcctac	Sacramento *et al.*, 1991
	N_1	Sequencing	tttgagacagcccctttg	Sacramento *et al.*, 1991
	NucRev	Sequencing	gacagttccgtcatctgc	Nel *et al.*, 1997
P gene	PHF	cDNA, PCR, sequencing	tat**gaattc**atcccaaacatgagcaa	This study
	PHR	PCR, sequencing	tta**gaattc**ggttagcaagatgtatagc	This study

Oligonucleotide sequences in **Bold** indicate EcoR1 restriction enzyme recognition sites, and the Underlined sequence indicates a BamH1 restriction enzyme recognition site.

2.2.4 Reverse transcription

For first strand cDNA synthesis of the N and P genes, total RNA (*ca.* 0.5 to 4.0 µg) was annealed with 12 pMol of the positive sense primers, N_3 and PHF (Table 2.1) respectively, at 65°C for 5 min. The reaction mixtures were cooled on ice and the RNA was reverse-transcribed at 37°C for 90 min in a 10 µl reaction mixture containing 50 U of Moloney Murine Leukemia Virus reverse transcriptase (M-MuLV) (USB^{TM}), 50 mM Tris-HCl (pH 8.3), 75 mM KCl, 3 mM $MgCl_2$, 10 mM DTT, 1 mM of each deoxynucleotide triphosphate (dNTP) and 5 U of RNasin ribonuclease inhibitor

(Promega). After completion of the reaction, the reaction mixtures were each diluted to a final volume of 50 µl with UHQ H_2O.

2.2.5 Polymerase chain reaction

The reaction mixtures (100 µl) contained 10 µl of diluted first strand cDNA, 100 µM of each dNTP, 1.5 mM $MgCl_2$, 50 mM KCl, 10 mM Tris-HCl (pH 9.0), 0.1% (v/v) Triton X-100 (supplied in *Taq* 10 × buffer) and 0.25 U of Taq DNA polymerase (Promega). For amplification of the N gene, 10 pMol of the JJ1 and N_3 primers were used, while for amplification of the P gene, 10 pMol of the PHF and PHR primers were used. The tubes were placed in a GeneAmp thermocycler (Model 2400; PE Applied Biosystems) for 30 cycles using the following temperature profile: denaturation at 94°C for 45 s, annealing at 50.5°C for 45 s, and elongation at 72°C for 90 s. The "hot start" protocol, with initial denaturation at 95°C for 5 min before addition of the enzyme, was carried out as described by Erlich *et al.* (1991).

2.2.6 Agarose gel electrophoresis

The PCR-amplified products were analyzed by agarose gel electrophoresis (Sambrook *et al.*, 1989). For this purpose, horizontal 0.8% (w/v) agarose gel slab gels were cast and electrophoresed at 100 V in 1 × TAE buffer (40 mM Tris-HCl, 20 mM Na.acetate, 1 mM EDTA; pH 8.5) using a Biorad Wide Mini Sub™ electrophoresis cell. The agarose gels were stained with ethidium bromide (0.5 µg/ml) and the DNA fragments were visualized by UV fluorescence. The DNA fragments were sized according to their migration in the gel as compared to that of a standard DNA molecular marker (Lambda DNA cut with EcoR1 and HindIII; Roche).

2.2.7 Purification of PCR products

The amplicons were purified from the agarose gel using a Geneclean™ kit (Bio 101 Inc.) according to the manufacturer's instructions. Briefly, the amplicons were excised from the agarose gel and mixed with *ca.* three volumes of a 3 M NaI solution. The agarose was dissolved at 55°C after which 5 µl of glassmilk® was added to the suspension. After incubation on ice for 5 min, the silica-bound DNA was pelleted by brief centrifugation,

and washed three times with ice-cold NEW Wash (a solution containing NaCl, Tris and EDTA in ethanol and water). The DNA was eluted from the silica matrix at 55°C for 2 - 3 min in a final volume of 12 µl of UHQ H_2O. A small aliquot of the eluate (1 µl) was electrophoresed on an agarose gel in order to determine the concentration of the DNA.

2.2.8 Cloning of the amplicons into the pGEM® T-Easy vector (Promega)

- *Ligation of DNA fragments*

Approximately 300 ng of purified amplicon and 50 ng of vector were ligated with 3 U of T4 DNA ligase and 1 µl of a 10 × ligation buffer (660 mM Tris-HCl, 10 mM DTT, 50 mM $MgCl_2$, 10 mM ATP; pH 7.5) at 4°C for 16 h in a final volume of 10 µl.

- *Preparation of competent cells*

The $CaCl_2$ method, as described by Sambrook *et al.* (1989), was used for the preparation of competent *E. coli* cells. A 1 ml overnight culture of *E. coli* JM109 cells was inoculated into 100 ml of preheated (37°C) sterile LB-broth, incubated at 37°C with agitation until an absorbency at 550 nm of 0.4 to 0.5 and then cooled on ice for 20 min. The cells from 30 ml of the culture were pelleted in Corex tubes by centrifugation at 4 000 rpm for 10 min at 4°C in a Sorvall® HB 4 rotor (Du Pont Instruments). The cells were suspended in 10 ml ice-cold 50 mM $CaCl_2$, incubated on ice for 1 h, pelleted as before and finally resuspended in 1 ml of the $CaCl_2$ solution. The competent cells were incubated on ice for at least 1 h before being used in transformation experiments.

- *Transformation of competent E. coli JM109 cells*

Competent cells were transformed using the heat-shock method as described by Sambrook *et al.* (1989). Competent cells (200 µl) were mixed with 5 µl of the respective ligation mixtures in sterile transformation tubes, and incubated on ice for 30 min. The cells were then incubated at 42°C for 90 s and rapidly chilled on ice for 2 min, after which 800 µl of pre-warmed LB-broth was added. After incubation at 37°C for 1 h with agitation, the recovered cells were plated in aliquots of 100 to 200 µl onto the surface of LB-agar plates supplemented with 100 µg/ml ampicillin. To allow for blue-white colour

selection, based on insertional inactivation of the *lac Z* gene, 40 µl of X-gal (2% (v/v) stock solution) and 10 µl of IPTG (100 mM stock solution) were also spread over the surface of the agar plates. The plates were then incubated overnight at 37°C and observed for the presence of recombinant transformants with a Gal$^-$ phenotype. These were selected for further characterization and grown overnight in 3 ml of LB-broth, supplemented with ampicillin.

2.2.9 Plasmid DNA extraction

Recombinant plasmids were isolated using the alkaline lysis method as described by Birnboim and Doly (1979). The cells from 1.5 ml of the overnight cultures were harvested by centrifugation at 15 000 rpm for 1 min in a microfuge tube using a benchtop microfuge. The bacterial pellets were suspended in 100 µl of Solution 1 (50 mM glucose, 25 mM Tris-HCl (pH 8.0), 10 mM EDTA) and incubated on ice for 5 min. The resultant spheroplasts were lysed following addition of 200 µl freshly prepared Solution 2 (1% SDS, 0.2 M NaOH). The mixtures were neutralized with the addition of 150 µl of Solution 3 (3 M sodium acetate), resulting in the precipitation of chromosomal DNA, high molecular weight RNA and proteins. After incubation on ice for 15 min, the precipitate was removed by centrifugation at 15 000 rpm for 5 min. The plasmid DNA was precipitated from the recovered supernatants by the addition of 2.5 volumes of ethanol. After incubation at -20°C for 1 h, the DNA was pelleted by centrifugation and washed twice with 70% ethanol. The vacuum-dried DNA pellets were resuspended in 50 µl 1 × TE buffer (10 mM Tris, 1 mM EDTA; pH 7.4).

2.2.10 Purification of recombinant plasmid DNA

Plasmid DNA (50 µl) was incubated with 1 µl RNase A (10 ng/µl) at 37°C for 30 min and then precipitated with 30 µl of a PEG/NaCl solution (20% (w/v) PEG 6000, 2.5 M NaCl). After incubation on ice for 30 min, the DNA pellet was obtained by centrifugation at 15 000 rpm for 20 min, washed twice with 70% ethanol, vacuum-dried and resuspended in 25 µl 1 × TE. This method yielded highly purified plasmid DNA and

strongly selects for the super-coiled covalently closed circular (ccc) form of plasmid DNA.

2.2.11 Restriction endonuclease digestion

Purified plasmid DNA (1 µg) was typically digested with 5 U of restriction enzyme (Roche) in the appropriate concentration salt (using the 10 × buffer supplied by the manufacturer). Reaction volumes were typically 15 µl, and incubation was carried out at 37°C for 90 min, after which the reaction was terminated by inactivation of the enzyme at 65°C for 3 min. Following restriction of the plasmid DNA, the restriction fragments were analyzed by agarose gel electrophoresis as described in Section 2.2.6.

2.2.12 Nucleotide sequencing

Sequencing of the cloned insert DNA was performed using an ABI PRISM™ Big Dye™ Terminator Cycle Sequencing Ready Reaction Kit (PE Applied Biosystems, Perkin-Elmer). Each reaction mixture contained 50 – 100 ng of PEG/NaCl-purified template DNA, 3.2 pMol primer, 2 µl Terminator Ready Reaction Mix, 2 µl of 5 × Buffer and UHQ water in a final reaction volume of 10 µl. Cycle sequencing was performed using the following program for 25 cycles: denaturation at 96°C for 30 s, annealing at 50°C for 15 s and extension at 60°C for 4 min. Following brief centrifugation, the extension products were precipitated by the addition of 8 µl UHQ water and 32 µl 95% ethanol (final concentration of 60%). The tubes were incubated at room temperature for 15 min, centrifuged at 15 000 rpm for 15 min and the supernatants carefully aspirated. The pellets were washed twice with 50 µl 70% ethanol, vacuum-dried for 10 - 15 min and then stored at -20°C. Prior to electrophoresis, the purified extention products were resuspended in 3.5 µl Blue dextran/EDTA loading buffer (PE Applied Biosystems), denatured for 2 min at 90°C and loaded onto the ABI Prism model 377 DNA sequencer.

2.2.13 Sequence analysis

The obtained nucleic acid sequences were edited with the ABI Prism Sequencing Analysis 3.1 and the ABI Prism Sequencing Navigator 1.0.1 software programs (PE

Applied Biosystems). The ClustalX program was used to perform multiple sequence alignment analyses of the South African viverrid rabies virus strain and several other rabies virus strains. The nucleic acid and amino acid sequences used in the alignment were extracted from GenBank[1] and had the following accession numbers: SAD, M31046; Pasteur virus, X03673; ERA, X55727; Pitman-Moore, X55729; RC-HL, AB009663; and Nishigahara, AB044824. The nucleic acid sequences were translated using Seqpup, and resulting amino acid sequences were used in PROSITE to identify possible protein kinase recognition sites. The sequence data is presented in multiple alignment formats generated by ClustalX, saved in postscript format and displayed in CorelDraw v. 9.0.

2.3 Results

2.3.1 cDNA synthesis of the N and P genes

Brain material of suspected rabid *Cynictis penicillata* (Yellow mongoose) were screened for rabies virus by means of fluorescence antibody typing (FAT) using monoclonal antibodies directed against the nucleoprotein. The viruses were serotyped further to distinguish between Canid and Viverrid serotypes. A viverrid rabies virus isolate (isolate 710/90) was identified and sequence analysis of the nucleoprotein and glycoprotein genes confirmed the viverrid biotype of the isolate (Olivier, 1997). This isolate was selected for further use in this investigation.

Following propagation of rabies virus isolate 710/90 in suckling mice, total RNA was extracted from the lyophilized brain material and used in reverse transcriptase-mediated PCR amplification of full-length cDNA copies of the N and P genes. To aid in the design of oligonucleotide primers for the specific amplification of the P gene, nucleic acid sequences of the P gene of ERA, PM and CVS-11 rabies virus strains (Larson and Wunner, 1990) as well as the Pasteur virus strain (Tordo *et al.*, 1986) were aligned using the ClustalX multiple sequence alignment program (Thompson *et al.*, 1997). Primers for the specific amplification of the full-length N gene were designed based on the

[1] The GenBank database website address: www.ncbi.nlm.nih.gov/GenBank/

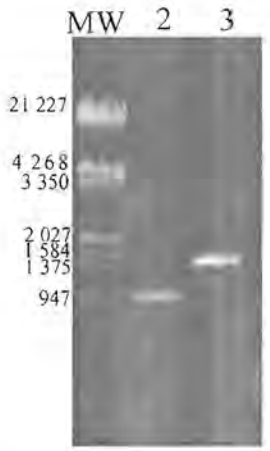


Fig. 2.1 Agarose gel electrophoretic analyses of the RT-PCR cDNA amplicons of viverrid rabies virus isolate 710/90. Lane MW, Molecular weight marker; Lane 2, PHF-PHR amplicon of the P gene; Lane 3, JJ1-N_3 amplicon of the N gene. The sizes of molecular weight marker, phage lambda DNA restricted with EcoR1 and HindIII, are indicated to the left of the figure in base pairs.

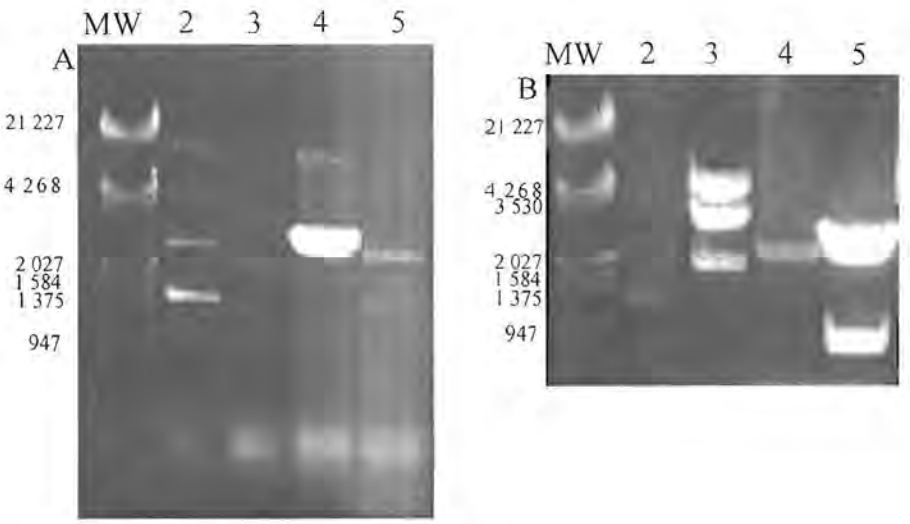


Fig. 2.2 Agarose gel electrophoretic analysis of the recombinant plasmids (A) NpGEM and (B) PpGEM, constructed by cloning of N and P gene cDNA into the pGEM-T Easy vector. (A) Lanes: MW, molecular weight marker; 2, uncut pGEM-T Easy; 3, uncut recombinant NpGEM; 4, EcoR1 and BamH1-digested pGEM-T Easy; 5, EcoR1 and BamH-digested recombinant NpGEM. (B) Lanes: MW, molecular weight marker; 2, uncut pGEM-T Easy; 3, uncut recombinant PpGEM; 4, EcoR1-digested pGEM-T Easy; 5, EcoR1-digested recombinant PpGEM. The sizes of molecular weight marker, phage lambda DNA restricted with EcoR1 and HindIII are indicated on the left of the figures in base pairs.

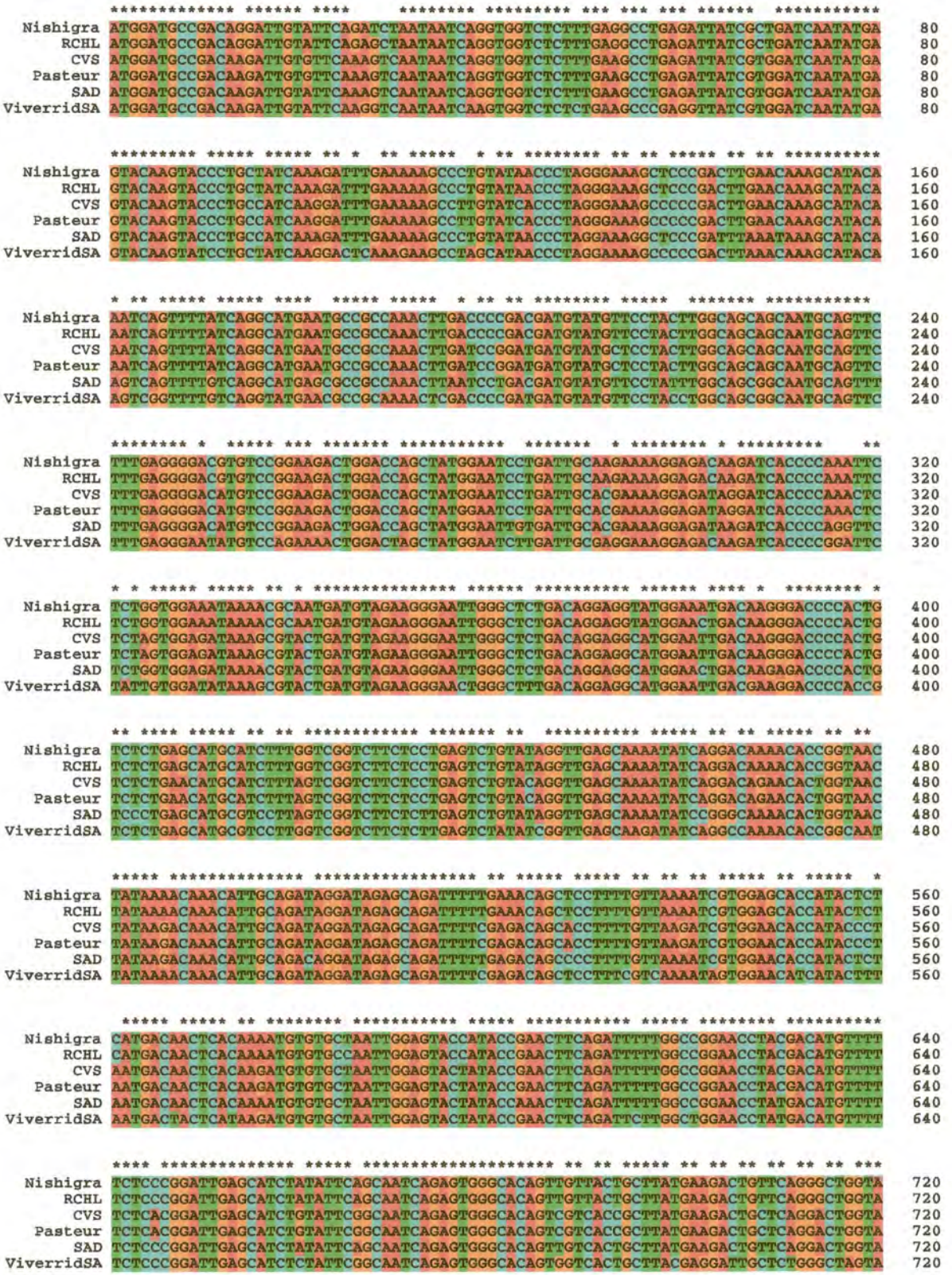
unpublished partial nucleotide sequence of a N gene of a viverrid rabies virus (Olivier, 1997). In order to facilitate cloning of PCR-amplified N and P genes, unique restriction endonuclease recognition sites were included in the primers. The oligonucleotide primer sequences and applications are listed in Table 2.1. Following agarose gel electrophoresis of the reaction mixtures, amplicons of the expected sizes were obtained when compared to the size of the DNA molecular weight marker. These corresponded to 1.4 kb for the full-length N gene (Fig. 2.1, lane 3) and *ca.* 900 bp for the full-length P gene (Fig. 2.1, lane 2).

2.3.2 Construction of recombinant pGEM® T-Easy vectors containing the full-length N and P genes

The use of *Taq* polymerase lacking a proofreading ability often result in amplicons containing a single non-template specific deoxyadenosine at the 5' ends of the amplified fragments. The pGEM®-T Easy Vector contains single 3'-thymidine overhangs at the insertion site, which results in a highly effective PCR-cloning system. Following gel-purification of the amplicons and ligation into pGEM®-T Easy, competent *E. coli* JM109 cells were transformed and colonies with a white phenotype were selected and cultured in LB-broth supplemented with ampicillin. Following plasmid DNA extractions, the plasmid DNA was analyzed for N and P gene-specific inserts by using restriction endonucleases of which the recognition sites have been incorporated during the design of the primers. The putative recombinant NpGEM clones were digested with BamH1 and EcoR1 and resulted in excision of a 1.4 kb DNA fragment, corresponding to the size of the expected full-length N gene (Fig. 2.2A, lane 5). Digestion of the selected putative recombinant PpGEM clones with EcoR1 resulted in excision of a DNA fragment of *ca.* 900 bp (Fig. 2.2B, lane 5). Recombinant plasmid DNA containing either the N or P gene, was selected and used for further DNA manipulations.

2.3.3 Sequencing of the viverrid rabies virus N and P genes

To verify the integrity of the cloned N and P genes, nucleotide sequences were determined by automated sequencing as described in Section 2.2.12. In addition to using the universal pUC/M13 forward and reverse sequencing primers, internal gene-specific


Nishigra
RCHL
CVS
Pasteur
SAD
ViverridSA

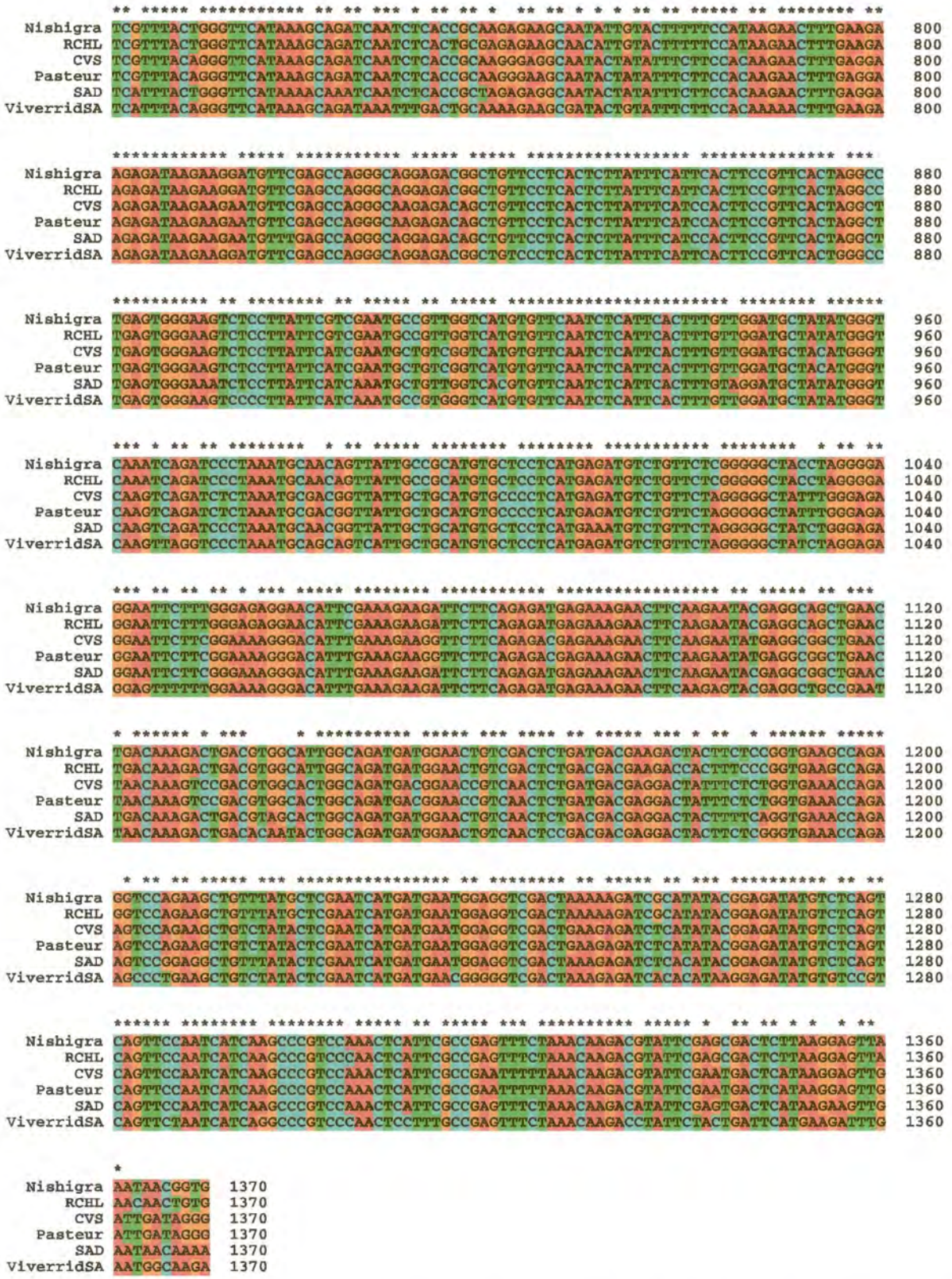
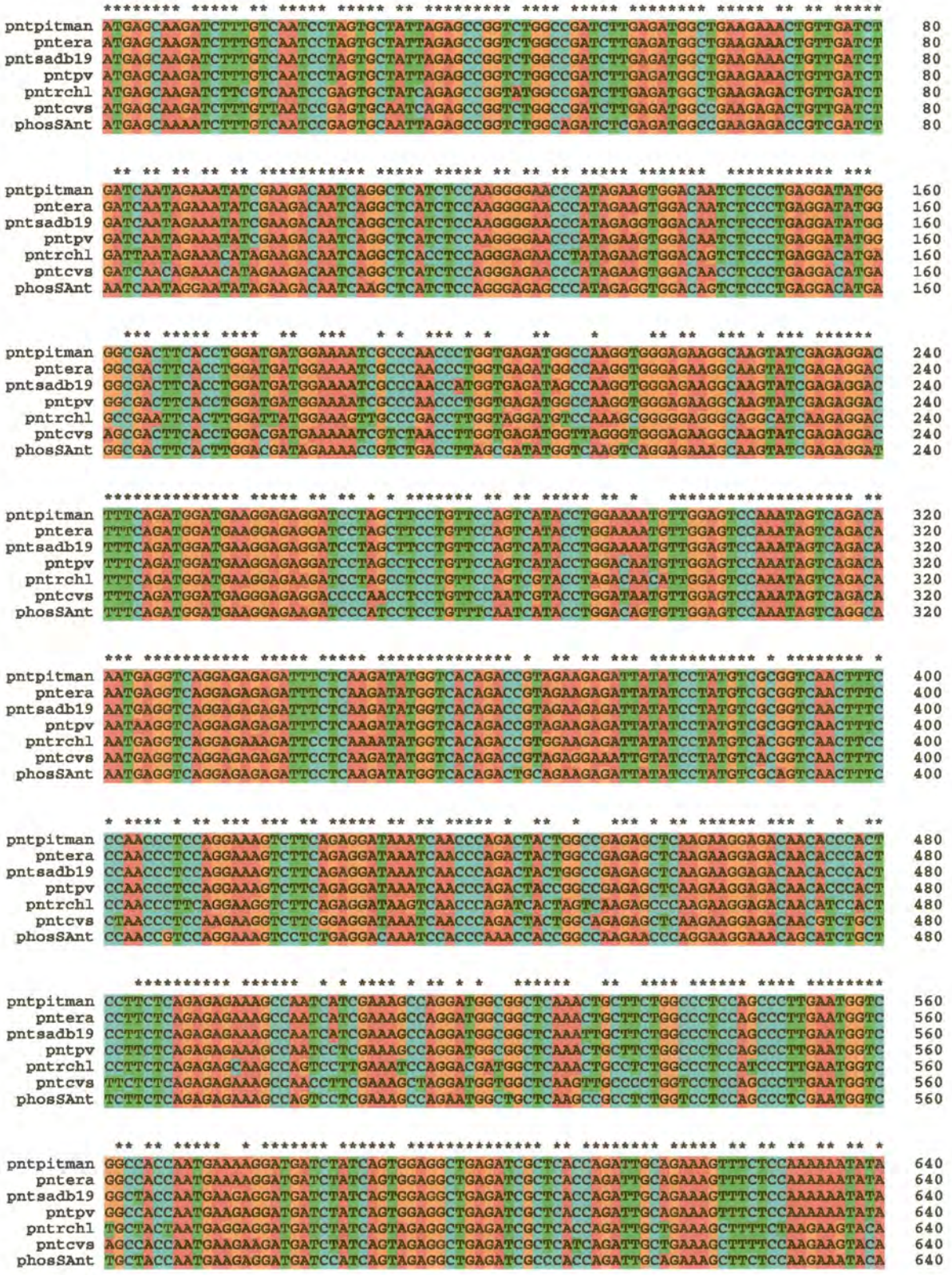


Fig. 2.3A Multiple sequence alignment of the 1 370 bp N gene of the South African viverrid rabies virus strain with other rabies virus strains using the ClustalX program, as described in Materials and Methods. GenBank accession numbers of rabies virus isolates used in this analysis: SAD, M31046; Pasteur virus, X03673; RC-HL, AB009663; Nishigahara, AB044824; CVS, X55727.


pntpitman
pntera
pntsadb19
pntpv
pntrchl
pntcvs
phosSAnt

```
          ******* ** ** ** ***** *** ******************  **** ***** ******** *******  **
pntpitman AGTTTCCCTCTCGATCCTCAGGGATACTCTTGTATAATTTTGAGCAATTGAAAATGAACCTTGATGATATAGTTAAAGAG  720
   pntera AGTTTCCCTCTCGATCCTCAGGGATACTCTTGTATAATTTTGAGCAATTGAAAATGAACCTTGATGATATAGTTAAAGAG  720
pntsadb19 AGTTTCCCTCTCGATCCTCAGGGATACTCTTGTATAATTTTGAGCAATTGAAAATGAACCTTGATGATATAGTTAAAGAG  720
    pntpv AGTTTCCCTCTCGATCCTCAGGGATACTCTTGTATAATTTTGAGCAATTGAAAATGAACCTTGATGATATAGTTAAAGAG  720
  pntrchl AGTTTCCTTCTCGGTCTTCAGGGATATTCTTGTATAATTTTGAGCAGTTGAAAATGAACCTTGATGATATAGTTAAGGAA  720
   pntcvs AGTTTCCCTCCCGATCTTCAGGAATATTCTTGTATAATTTTGAGCAACTGAAGATGAACCTTGATGACATAGTTAAAGAG  720
 phosSAnt AGTTTCCCTCTCGATCATCAGGAATATTCTTGTATAATTTTGAGCAGTTGAAAATGAATCTTGATGACATAGTTAGAGAG  720

          *********** ** ** ** ****** * ** * *** ** ****** **** ** ***** ** ***** ********
pntpitman GCAAAAAATGTACCAGGTGTGACCCGTTTAGCCCATGACGGGTCCAAACTCCCCCTAAGATGTGTACTGGGATGGGTCGC  800
   pntera GCAAAAAATGTACCAGGTGTGACCCGTTTAGCCCATGACGGGTCCAAACTCCCCCTAAGATGTGTACTGGGATGGGTCGC  800
pntsadb19 GCAAAAAATGTACCAGGTGTGACCCGTTTAGCCCATGACGGGTCCAAACTCCCCCTAAGATGTGTACTGGGATGGGTCGC  800
    pntpv GCAAAAAATGTACCAGGTGTGACCCGTTTAGCCCGTGACGGGTCCAAACTCCCCCTAAGATGTGTACTGGGATGGGTCGC  800
  pntrchl GCAAAAAATGTGCCAGGCGTAACCCGTTTGGCCCATGACGGATCCAAACTCCCCCTAAGATGTGTACTGGGATGGGTCGC  800
   pntcvs GCAAAAAATGTACCGGGCGTGACCCGTCTGGCCCATGATGGATCCAAAATCCCCCTGAGATGCGTACTGGGATGGGTCGC  800
 phosSAnt GCAAAAAATGTGCCAGGTGTGACCCGTTTGGCACATGACGGATCCAAACTCCCTCTGAGATGTGTGCTGGGGTGGGTCGC  800

            * ** ** ** ** ******** **  * *****  *  ******* *  ** ** ***********  * *******
pntpitman TTTGGCCAACTCTAAGAAATTCCAGTTGTTAGTCGAATCCGACAAGCTGAGTAAAATCATGCAAGATGACTTGAATCGCT  880
   pntera TTTGGCCAACTCTAAGAAATTCCAGTTGTTAGTCGAATCCGACAAGCTGAGTAAAATCATGCAAGATGACTTGAATCGCT  880
pntsadb19 TTTGGCCAACTCTAAGAAATTCCAGTTGTTAGTCGAATCCGACAAGCTGAGTAAAATCATGCAAGATGACTTGAATCGCT  880
    pntpv CTTGGCCAACTCTAAGAAATTCCAGTTGTTAGTCGAATCCAACAAGCTGAGTAAAATCATGCAAGATGACTTGAATCGCT  880
  pntrchl TCTGGCCAATTCTAAGAAATTCCAATTGTTAGTCGAAGCCAACAAGCTAAATAAAATCATGCAAGATGATTTGAATCGCT  880
   pntcvs TTTGGCTAATTCCAAAAAATTCCAATTACTAGTCGAGGCTGACAAGCTAAGCAAAATCATGCAAGATGATTTGAATCGCT  880
 phosSAnt CCTAGCCAACTCTAAGAAATTCCAGTTGTTGGTCGAGTCTGACAAGCTAAGCAAGATTATGCAAGATGATCTAAATCGCT  880

          ** **** ******
pntpitman ATACATCTTGCTAA   894
   pntera ATACATCTTGCTAA   894
pntsadb19 ATACATCTTGCTAA   894
    pntpv ATACATCTTGCTAA   894
  pntrchl ATGCATCTTGCTAA   894
   pntcvs ATACATCCTGCTAA   894
 phosSAnt ATACATCTTGCTAA   894
```

Fig. 2.3B Multiple sequence alignment of the 894 bp P gene of the South African viverrid rabies virus strain with other rabies virus strains using the ClustalX program. GenBank accession numbers of the isolates used: Pitman-Moore, X55729; ERA, X55728; SAD-B19, M31046; Pasteur virus (PV), X03673; RC-HL, AB009663; CVS, X55727. The rabies virus strains in the alignment is indicated with the following names: Pitman-Moore, pntpitman; ERA, pntera; SAD-B19, pntsadb19; Pasteur virus, pntpv; RC-HL, pntrchl; CVS, pntcvs and the South African viverrid strain is named phosSAnt.

primers were also designed (Table 2.1) to obtain the complete nucleic acid sequence of the respective genes. The obtained nucleic acid sequence for the N gene is indicated in Fig. 2.3A and that for the P gene in Fig. 2.3B.

The N gene of the South African viverrid rabies virus strain was found to be 1 370 nucleotides in length. Comparison of the sequence to other rabies virus strains indicated 93.79% homology. The deduced amino acid sequence of the N protein encoded by the viverrid N gene indicated that the protein consists of 450 amino acid residues and has an estimated molecular weight of 55 kDa. Alignment comparison of the amino acid sequences of the viverrid virus strain indicated 96.2% identity to the N protein of the CVS rabies virus strain (Fig 2.4A). The C-terminal 74 amino acids of the CVS rabies virus strain N protein have been reported to be involved in binding to the P protein (Schoehn *et al.*, 2001). Comparison of this region, between the viverrid and CVS strains indicated a 96% identity (Fig 2.4A). The sequence spanning amino acids 298-352 has been reported to be involved in binding of the Pasteur virus strain N protein to the viral ssRNA genome (Kouznetzoff *et al.*, 1998). Comparison of this region to the corresponding region of the viverrid rabies virus included in the analysis indicated 98% identity. Only a single amino acid change in position 328 was noted (threonine in PV to an alanine in the viverrid strain).

The phosphoprotein gene of the South African viverrid rabies virus strain is 894 nucleotides in length, displays 90.4% and 90.5% identity with the CVS and Pasteur virus strains, respectively. The encoded P protein is composed of 297 amino acids, with a predicted molecular weight of 36 kDa, and displays 90% identity to both the Pasteur virus and the CVS rabies virus strains (Fig. 2.4B). The N-terminal 19 amino acids, which is the region responsible for binding to the L protein (Chenik *et al.*, 1998), was shown to be conserved in all the strains used in this comparison. This high level in sequence conservation confirms the functional importance of this domain (Chenik *et al.*, 1998).

The amino acid sequences of the N and P proteins of the viverrid rabies virus strain were also investigated in order to identify sequences characteristic of phosphorylation sites.

The consensus phosphorylation sites for protein kinase C and casein kinase II are summarized in Table 2.2. The only other kinase recognition sites on the N protein was for Tyrosine kinase at amino acids 359 – 367 (RDEKELQEY) and 398 – 404 (RSPEAVY). These sites are not conserved in the Nishighara and the RC-HL rabies virus strains and were not considered during any further analysis. Phosphorylation of these sites will be determined by various factors, such as availability of the kinase within the host cell, and the presentation of the site within the protein for recognition by the kinase. Phosphorylation of a protein by a specific kinase do not guarantee any functional significance for the protein, and will only be known once the function of the phosphorylated and unphosphorylated forms of the protein is compared.

Table 2.2. Consensus protein kinase recognition sites identified in the viverrid rabies virus N and P proteins.

Rabies virus protein	**Kinase**	**Phosphorylation site**	
		Location	**sequence**
Nucleoprotein	Protein Kinase C	16-18	SLK
		190-192	THK
		295-297	SGK
	Casein Kinase II	107-110	SIVD
		114-117	TDVE
		133-136	TVSE
		215-218	SRIE
		231-234	TAYE
		252-255	TAKE
		375-378	TKTD
		389-392	SDDE
		437-440	SFAE
Phosphoprotein	Protein Kinase C	162-164	SQR
		168-170	SSK
		210-212	SKK
		271-273	SKK
	Casein Kinase II	95-98	SYLD
		121-124	TAEE
		140-143	SSED
		162-165	SQRE
		189-192	TNEE

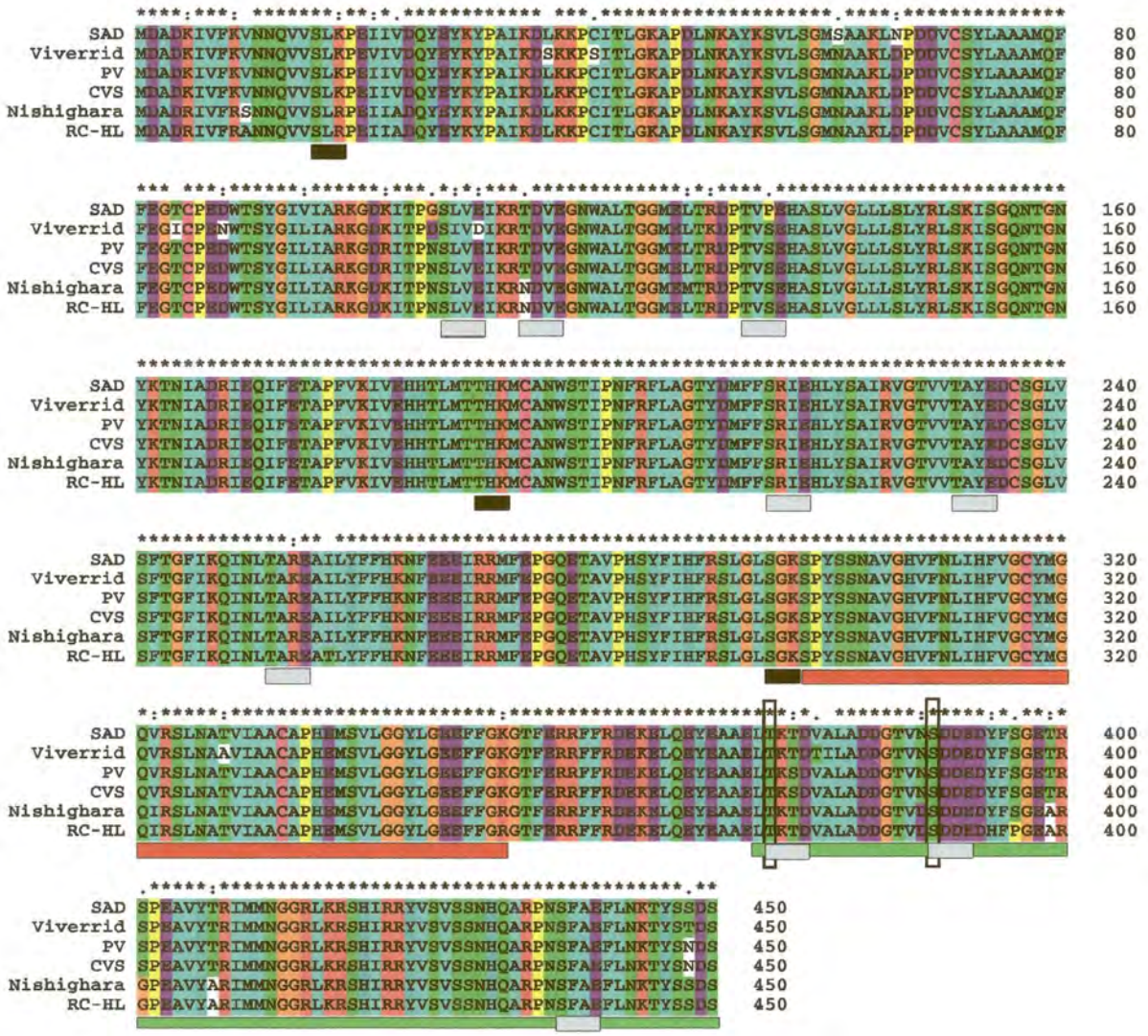


Fig. 2.4A Multiple alignment of the amino acid sequences of the N proteins of different rabies virus strains, generated by the ClustalX program, as described under Materials and Methods. The phosphorylated serine at position 389 and threonine at position 375 are indicated in the block. The black and gray bars under the alignmets indicate the kinase recognition sites for Protein Kinase C and Casein Kinase II, respectively, as predicted by PROSITE. The red bar indicate the ssRNA-binding domain of the rabies virus N protein, and the green bar

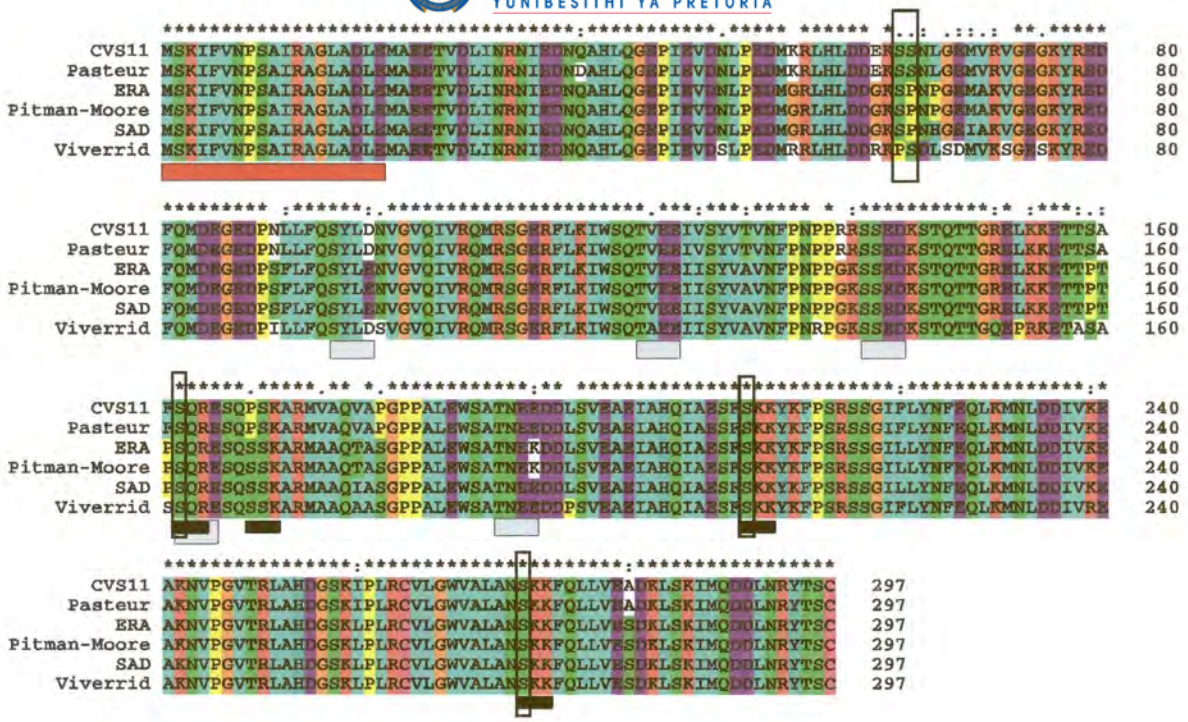


Fig 2.4B Multiple alignments of the amino acid sequences of the P proteins of different rabies virus strains, generated by the ClustalX program, as described in Materials and Methods. The serine residues phosphorylated by the Rabies Virus Protein Kinase at position 63 and 64 are indicated by a solid black box. The serine residues phosphorylated by Protein Kinase C at position 162, 210, 271 are also boxed black. The black bars under the alignments indicate the Protein Kinase C recognition sites as predicted by PROSITE, while the grey bars indicate the Casein Kinase II consensus sequences as predicted by PROSITE. The red bar indicate the N-terminal 19 amino acids involved in P protein binding with the L protein.

2.4 Discussion

In this chapter, the full-length nucleo- and phosphoprotein genes of a South African viverrid rabies virus isolate were cloned and sequenced. The nucleotide sequences were translated and compared to the cognate proteins of other rabies viruses. This was aimed at characterizing structurally important domains of these two proteins and to provide information regarding possible consensus phosphorylation sites.

Functional domains on the viverrid N protein, previously identified for other rabies virus strains, were compared. Amino acid residues 298 to 352 of the Pasteur virus strain is crucial for ssRNA binding and the corresponding amino acids of the viverrid strain showed only one amino acid change at position 328 (threonine in PV and alanine in viverrid. Since threonine is a neutral polar amino acid, therefore uncharged under normal physiological conditions, and alanine is a nonpolar amino acid, there would be minimal structural differences between the N proteins of these two strains and very little difference in their RNA-binding ability. The C-terminal 74 amino acid residues play an important role during complex formation between the N and the P protein. Within this domain, we identified three amino acid differences, of which the substitution of a non-polar Val-379 in CVS with a polar threonine amino acid in the viverrid strain can effect the complex formation between these two recombinant proteins. The N protein amino acid sequence of the Pasteur rabies virus strain has previously been compared with that of VSV (Indiana serotype) (Tordo *et al.*, 1986a). This study reported conserved segments of amino acids between the N proteins of these two strains, and specifically at amino acids 72 to 112 with 41.5% homology, 140 to 150 with 54.5% homology, 225 to 247 with 39% homology and 268 to 302 with 48.5% homology.

The deduced amino acid sequences for both the P and the N proteins were used to identify consensus protein kinase recognition sites that can potentially be recognized and phosphorylated by these kinases. Three of the four identified PKC recognition sites in the phosphoprotein corresponds to known phospho-acceptors (Gupta *et al.*, 2000): *viz* Ser-162, Ser-210 and Ser-271. Although PKC has been reported to phosphorylate the rabies virus phosphoprotein, the known phospho-acceptors of the nucleoprotein, Ser-389

and Thr-375, do not correspond to the identified PKC recognition sites (Gupta *et al.*, 2000). Anzai *et al.* (1997) identified these phospho-amino acids with a monoclonal antibody directed against a specific phosphatase-sensitive epitope as well as through amino acid hydrolysis. It is very likely that there are other phosphorylated amino acids besides Ser-389 and Thr-375 in the Nucleoprotein, with other kinases involved in phosphorylating the rabies virus N protein.

With the use of the PROSITE database and software, nine Casein kinase II consensus sequences in the N protein amino acid sequence were identified. Two of these corresponded to the location of the known phospho-acceptors, amino acid residues 375-378 (TKTD) and 389-392 (SDDE). Gupta *et al.* (2000) reported that, in addition to several isomers of PKC, the rabies virus P protein are also phosphorylated by a unique heparin-sensitive non-CKII protein kinase. This enzyme appears to be selectively packaged within the matured virions, and were designated RVPK (rabies virus protein kinase). The authors were unable to identify this protein kinase by comparing its biochemical properties with those of other known cellular protein kinases in protein data banks. This work by Gupta *et al.* (2000) open the door for further studies on the phosphorylation of rabies virus proteins and its role during virus transcription and replication.