MOLECULAR CHARACTERIZATION OF SEVERE ACUTE RESPIRATORY SYNDROME (SARS) CORONAVIRUS - NUCLEOCAPSID PROTEIN

by

VINITA SINGH CHAUHAN

B.Sc, Mumbai University, 1998
M.Sc, Mumbai University, 2000

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Diagnostic Medicine and Pathobiology

College of Veterinary Medicine

KANSAS STATE UNIVERSITY

Manhattan, Kansas

2006

ABSTRACT

Severe acute respiratory syndrome (SARS) is caused by an enveloped, positive-stranded RNA virus, the SARS coronavirus (SARS-CoV). Coronaviruses along with the arteriviruses are placed in the order, *Nidovirales*. Even though nidovirus replication is restricted to the cytoplasm, the nucleocapsid protein (N) of several coronaviruses and arteriviruses, localize to the nucleolus during infection. Confocal microscopy of N protein localization in Vero cells infected with the SARS-CoV or transfected with the SARS-CoV N gene failed to show presence of N in the nucleoplasm or nucleolus. Recombinant N remained cytoplasmic after the addition of leptomycin B (LMB), a drug that inhibits nuclear export. SARS-CoV N possesses a unique lysine-rich domain, located between amino acids 369-389, which possesses several nuclear localization signal (NLS) and nucleolar localization signal (NoLS) motifs. A chimeric protein composed of the 369-389 peptide substituted for the NLS of equine infectious anemia virus (EIAV) Rev protein (ERev) showed no nuclear localization activity. Three negatively charged amino acids, located at positions 372, 377 and 379 in SARS-CoV N were hypothesized to play a role in the loss of nuclear targeting. Substitution of aspartic acid-372 with alanine restored nuclear localization to the chimeric protein. A full-length recombinant SARS-N protein with the alanine-372 substitution localized to the nucleus. Therefore, the presence of an aspartic acid at position 372 is sufficient to retain N in the cytoplasm The mechanistic basis for how aspartic acid-372 interrupts nuclear transport is unknown, but may lie in the electrostatic repulsion with negatively charged amino acids located within the NLS binding pocket of importin-alpha.

MOLECULAR CHARACTERIZATION OF SEVERE ACUTE RESPIRATORY SYNDROME (SARS) CORONAVIRUS - NUCLEOCAPSID PROTEIN

by

VINITA SINGH CHAUHAN

B.Sc, Mumbai University, 1998
M.Sc, Mumbai University, 2000

A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Diagnostic Medicine and Pathobiology

College of Veterinary Medicine

KANSAS STATE UNIVERSITY

Manhattan, Kansas

2006

Approved by :

Major Professor
Dr. Raymond R. Rowland

ABSTRACT

Severe acute respiratory syndrome (SARS) is caused by an enveloped, positive-stranded RNA virus, the SARS coronavirus (SARS-CoV). Coronaviruses along with the arteriviruses are placed in the order, *Nidovirales*. Even though nidovirus replication is restricted to the cytoplasm, the nucleocapsid protein (N) of several coronaviruses and arteriviruses, localize to the nucleolus during infection. Confocal microscopy of N protein localization in Vero cells infected with the SARS-CoV or transfected with the SARS-CoV N gene failed to show presence of N in the nucleoplasm or nucleolus. Recombinant N remained cytoplasmic after the addition of leptomycin B (LMB), a drug that inhibits nuclear export. SARS-CoV N possesses a unique lysine-rich domain, located between amino acids 369-389, which possesses several nuclear localization signal (NLS) and nucleolar localization signal (NoLS) motifs. A chimeric protein composed of the 369-389 peptide substituted for the NLS of equine infectious anemia virus (EIAV) Rev protein (ERev) showed no nuclear localization activity. Three negatively charged amino acids, located at positions 372, 377 and 379 in SARS-CoV N were hypothesized to play a role in the loss of nuclear targeting. Substitution of aspartic acid-372 with alanine restored nuclear localization to the chimeric protein. A full-length recombinant SARS-N protein with the alanine-372 substitution localized to the nucleus. Therefore, the presence of an aspartic acid at position 372 is sufficient to retain N in the cytoplasm The mechanistic basis for how aspartic acid-372 interrupts nuclear transport is unknown, but may lie in the electrostatic repulsion with negatively charged amino acids located within the NLS binding pocket of importin-alpha.

TABLE OF CONTENTS

LIST OF FIGURES……ii
ACKNOWLEDGEMENTS..……..iii
DEDICATION……v

SECTION I : Introduction and literature review
Molecular Characteristics of Severe Acute Respiratory Syndrome……1
Introduction……..……....……..2
Table 1……..……17
References……..……19

SECTION II : Intracellular localization of the SARS coronavirus nucleocapsid protein: Absence of nucleolar accumulation during infection and after expression as a recombinant protein in Vero cells…….…….……26
Abstract……..……27
Introduction.....28
Results and discussion…….....31
References……..……42

SECTION III: An aspartic acid residue at position 372 of SARS-CoV N protein is critical for its cytoplasmic retention……...46
Abstract……..……47
Introduction.....48
Materials and methods…..……52
Results.....56
Discussion……..……66
References……..……67

SECTION IV: Conclusions and future directions…..……72

LIST OF FIGURES

SECTION I

Figure 1.1 Structure of SARS coronavirus……4
Figure 1.2 Genome organization of SARS genome…….……5
Figure 1.3 Sequence homology between SARS-N and group II coronavirus N protein..……....7
Figure 1.4 Diagrammatic representation of SARS-N domains ……..13

SECTION II

Figure 2.1 NLS motifs in SARS-CoV N protein ……...32
Figure 2.2 Distribution of N in SARS-CoV infected cells ……34
Figure 2.3 Localization of SARS-CoV N-EGFP in Vero cells ……36
Figure 2.4 Effect of LMB on localization of SARS-N-EGFP in live cells……39

SECTION III

Figure 3.1 NLS motifs in SARS-CoV N protein 369-389 peptide …….……54
Figure 3.2 Localization properties of EGFP-ERev-SARS-N chimeric protein …..... 57
Figure 3.3 Localization properties of mutated EGFP-SARS-N proteins ….……60
Figure 3.4 Intracellular distribution of EGFP fluorescence in live cells……..63
Figure 3.5 Intracellular localization of EGFP tagged SARS CoV N protein Fragments……65

ACKNOWLEDGMENTS

I would like to thank some of the people who made this work possible, feasible and delightful. I am extremely grateful to the faculty at the Department of Diagnostic Medicine and PathoBiology. Kansas State University –College of Veterinary Medicine. Special thanks to my major advisor Dr. Raymond Rowland for his patience and guidance through my scientific endeavors. Its his wisdom that has made this research and this dissertation possible. I also wish to extend my appreciation to the members of my graduate committee Dr. T.G. Nagaraja ; who has been the greatest source of encouragement, without his constant support, this dissertation would not have been complete. I also thank Dr. Derek Mosier for his invaluable input in the editing of my dissertation and Dr. Om Prakash for his scientific insight and guidance. I would also like to thank the Department Chair, Dr. M.M. Chengappa, for all his support.

My graduate studies would not have been the same without the friendship and input from my fellow graduate students; Luna Aguirre, Paula Schneider, Dal Young Kim and John Patton. I feel truly blessed to have friends like you and although I could never repay you enough for all you have done, I hope you understand how indebted I am. A special mention to the members of the Rowland Lab; your presence has made what could have been a truly difficult time, very memorable and enjoyable. I am extremely grateful.

Lastly, I would like to thank my family for their constant support, love and encouragement to pursue my interests; even when the interests went beyond the

boundaries of language and geography. My father, Dr. H.V.S Chauhan is my greatest inspiration-Thank you for believing in me when I did not believe in myself; my mother Manju Chauhan; who has prayed for this more than I have , I cannot thank her enough and my sister and partner in crime Neeta Chauhan for listening to all my complaints, insecurities and frustrations and through it all reassuring and encouraging me whole heartedly. I also wish to thank my husband and best friend Ram Gandhi; for his eternal optimism, endurance and humor; his companionship has made my journey through graduate school a wonderful time to remember. I am so thankful to all of you for these years that you have selflessly given and I have selfishly taken.

I would like to share the success of this research with every one of you. However, I alone remain responsible for errors and imperfections that may unintentionally have remained in this piece of work.

DEDICATION

I would like to dedicate this thesis to my parents Harsh and Manju Chauhan. There is no doubt in my mind that without their constant support and unconditional love, I could not have completed this research.

Section I

Introduction and Literature Review

Characteristics of Severe Acute Respiratory Syndrome- Coronavirus Nucleocapsid Protein

Introduction

Late in 2002, an outbreak of intense pneumonia infections emerged and spread in the GuangDong province of mainland China. An airborne pathogen was determined to be the cause of this illness, which was designated Severe Acute Respiratory Syndrome (SARS). A global warning was issued by the World Health Organization (WHO). Over 4,300 SARS cases were reported by late April of 2003, resulting in nearly 250 deaths. Most infections occurred after domestic exposure to another SARS patient. The epidemic eventually, spread to 30 different countries. By early 2004, over 8000 people were infected worldwide; including 800 who died from respiratory failure. Prompt implementation of diagnostic and preventative measures brought the spread of SARS infections under control by early 2004. The pathogen was determined to be a virus and the physical characteristics and genome organization led to the assignment of this virus in the Coronavirus family.

Taxonomy

Coronaviruses and Toroviruses are members of the *Coronaviridae* family, which along with the *Roniviridae* and *Arteriviridae* families form the order *Nidovirales*. Coronaviruses are enveloped, single stranded, positive sense RNA viruses that cause respiratory and enteric diseases in humans and animals. Coronaviruses were initially divided into four antigenic groups. However, differentiation based on the monoclonal antibody reactivity, host specificity and genome sequencing replaced the serological classification, and eventually classified the family into three groups. Groups 1 and 2

contain mammalian viruses, while group 3 primarily consists of avian viruses. Group 2 viruses are also characterized by the presence of a gene encoding the hemagglutinin esterase (HE) protein, which is absent in the other groups (33). Coronaviruses are highly host- specific and do not infect other organisms. However, the etiologic agent SARS coronavirus (SARS-CoV) (24) is capable of interspecies transmission. The original host of this virus was determined to be Chinese Horseshoe bats (*Rhinolopus sinicus*) and has since been isolated from a variety of animals including raccoon dogs, palm civets, Chinese ferret badgers and rats. (9, 16, 20). Human coronaviruses cause 30% of colds, most of them do not infect the lower respiratory tract like SARS-CoV infections.

SARS-CoV was initially believed to be a representative of a new fourth group in the Coronavirus family, due to lack of obvious similarities with any other established group. However, closer analysis of the replicase gene of the virus revealed that the virus is most closely related to the group 2 of the coronavirus family (35). Thus its placement in a new group, 2b, has been proposed.

Virus structure and genome organization

The viruses are named coronaviruses due to the appearance of the viral particles under a microscope. The presence of the spike protein that is present on the surface of the viral particles gives it a "crown-like" appearance (Fig 1.1). This protein is responsible for viral attachment and entry. The virus also has a number of other structural proteins, including

the Membrane glycoprotein (M), the Envelope protein (E) and the Nucleocapsid (N) protein as well as the single stranded RNA.

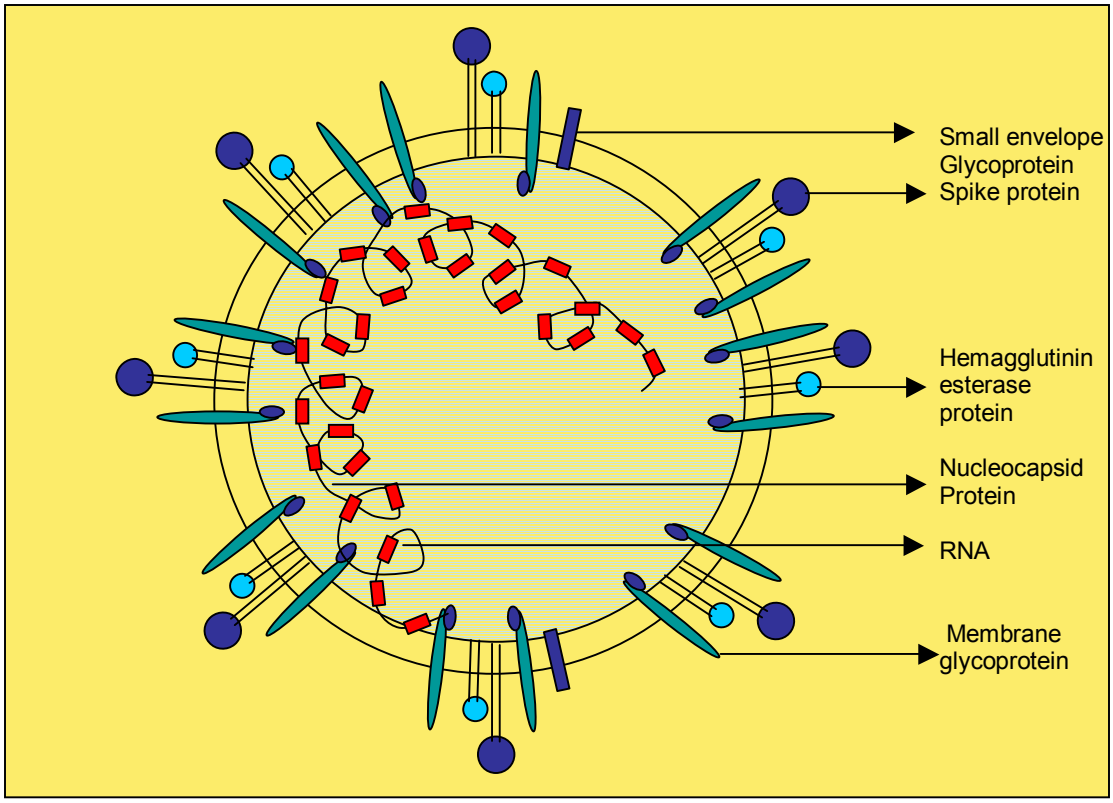


Fig 1.1 Structure of SARS coronavirus

SARS genome is approximately 29.7 kb and contains 40% GC residues. The genomic organization of the coding sequences of SARS-CoV is typical of other coronaviruses. The gene order is [5′-replicase (rep), spike(S)-envelope (E)-membrane (M) and nucleocapsid (N)-3′]. The rep gene comprises a large portion of the genome and codes for two polyproteins which undergo co-translational proteolytic processing. The genome also encodes multiple non-structural proteins (Fig 1.2; 33). Coronaviruses (CoV) replicate within the cytoplasm of the host cell. The positive-sense, single stranded genome contains a 5′ cap structure and a polyadenylated 3′ tail. The structural proteins

are translated from a 3´-coterminal nested set of subgenomic mRNAs. Each subgenomic RNA contains a common leader sequence identical to the 5´ leader sequence of the genome. The subgenomic mRNAs are formed by a discontinuous transcription mechanism during negative strand synthesis (Fig 1.2). Due to this discontinuous transcription mechanism, there is a high degree of RNA recombination.

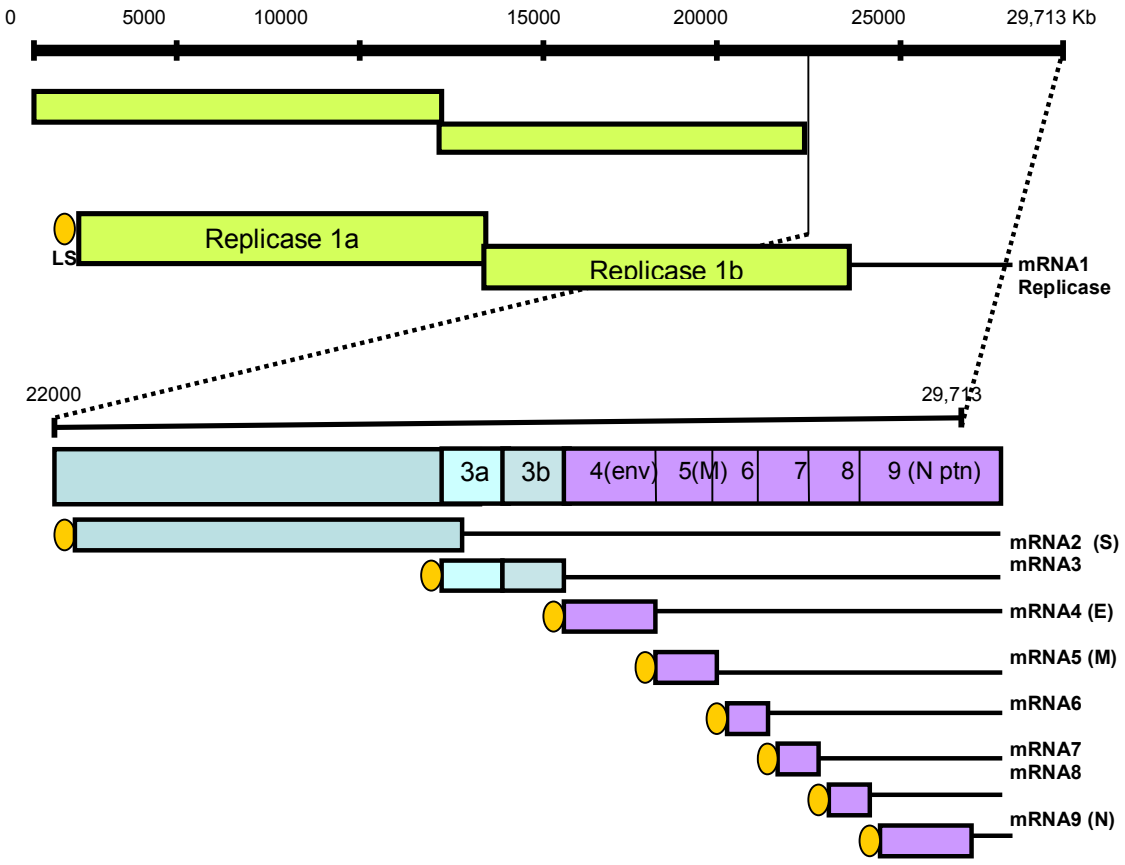


Fig 1.2 Genome organization of SARS genome.

Nucleocapsid (N protein):

The nucleocapsid protein is a major structural protein of the coronavirus. The SARS-CoV is 422 amino acids in length and 47KDa in mass. SARS-CoV n protein shares only 20-30% homology with N proteins from other known coronaviruses. PSORT sequence analysis reveals the presence of multiple nuclear localization signal sequences (NLSs) as well as presence of a unique leucine rich region and a serine-arginine rich region. SARS-N does not contain any cysteine residues and recombinant protein expression results in protein that is labile and easily degradable (6, 40, 41, 44)

Sequence homology of SARS-CoV N protein with other Coronavirus N proteins

Figure 1.3 displays the sequence alignment between SARS N with N proteins of two members of group 2 Bovine Coronavirus (BCV) and Murine hepatitis virus (MHV- A59). Surprisingly, even though SARS-N was sequentially most similar to this group, the percentage identity was is found to be very low (31% with MHV and 29% with BCV). When a sequence comparison was performed with members of the other groups of the family, the percentage identity was further decreased (21% with TGEV and PDEV and 26% with IBV; data not shown).

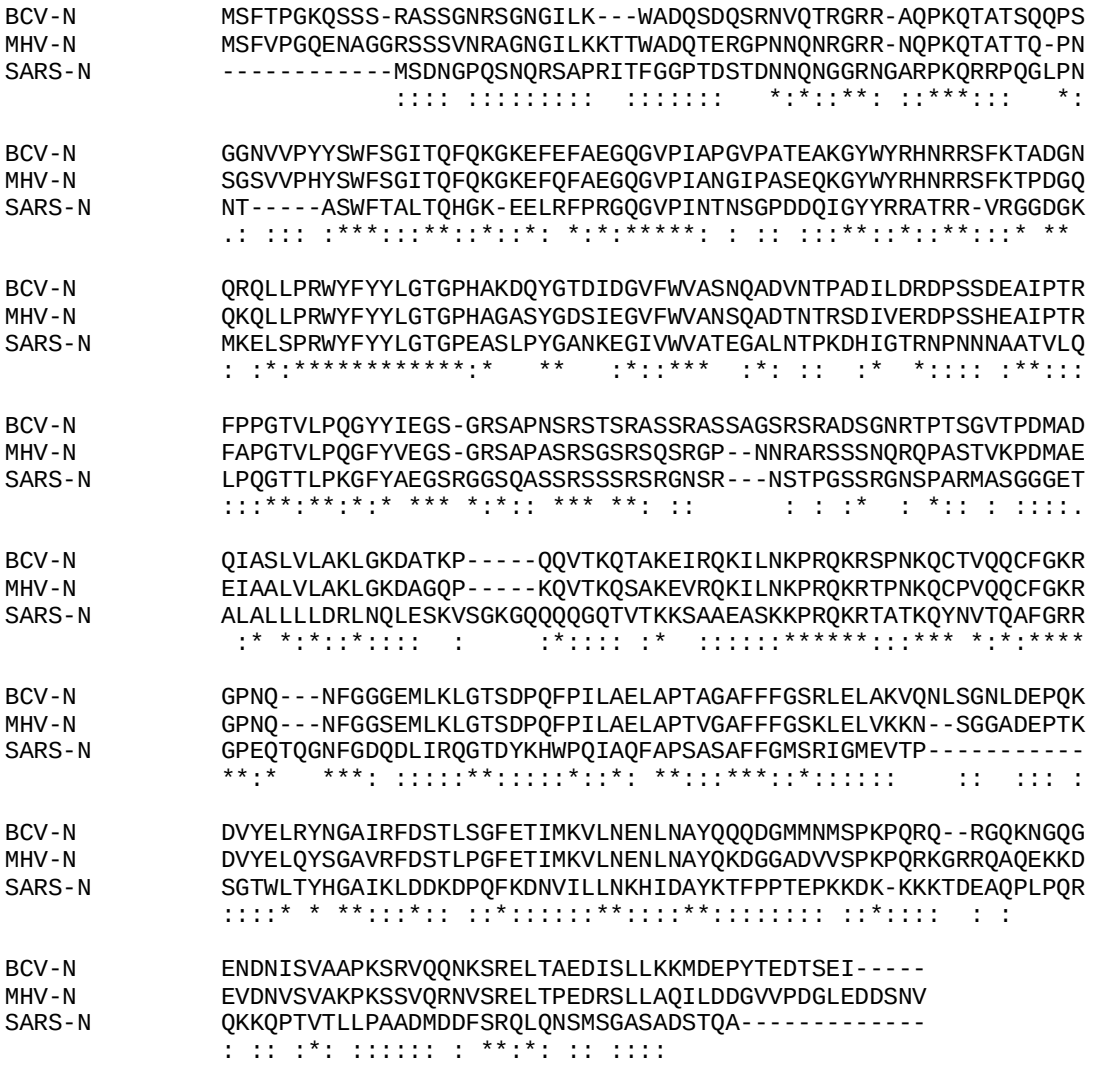


Fig 1.3 Sequence homology between SARS-N and group II coronavirus N proteins. Sequences from the nucleocapsid proteins from the coronaviruses of group II and SARS-CoV mentioned above were obtained from Genbank and multiple alignment of these sequences was done using CLUSTAL W analysis. The common amino acids are denoted by (*) and amino acid that are conserved in 2 out of the three sequences is denoted by (**:**).

Structure and functions of SARS-CoV N protein

Multiple roles have been attributed to the N protein of coronaviruses throughout the virus life cycle. Primarily, the association of N protein with genomic RNA forms the ribo-nucleocapsid structure and this in turn associates with the M protein to form the viral core. N protein plays a role in viral pathogenesis and anti-N monoclonal antibodies have been shown to protect mice from lethal infection as well as inhibit viral transcription. N protein is believed to play a role in transcription of viral genome by its association with the viral genomic and sub genomic RNAs (1, 31, 34). This interaction is mediated by the binding signals in the leader RNA sequences and is specific to the viral RNA.

The Coronavirus N protein can be divided into three putative regions, the N-terminal domain, the RNA-binding domain and a C-terminal domain. The N and C terminal are believed to play a role in molecular interactions with other proteins. The C-terminal end plays an important role in dimerization of the protein. However, the mid portion of SARS-N interacts with the M protein and hnRNP-A1 protein and the N-terminal domain (between amino acids 45-181) of the protein is shown to interact with viral RNA (7, 10, 13, 15, 28). Therefore, a recent study by Chang et al. (2) analyzed the modular organization of SARS-CoV nucleocapsid protein. Using NMR (nuclear magnetic resonance) techniques their observations proved that SARS-CoV N protein consists of two non-interacting, independent structural domains. The first domain is the N-terminal RNA interacting domain whereas the residues of the second domain at position 248-365

at the C-terminal are termed the dimerization domain. These domains are surrounded by flexible linkers and these linkers act as potential interaction sites with other proteins.

The most important biochemical property of the N protein is its capacity to self-associate. N proteins of many coronaviruses can form dimers and oligomers. Dimer formation is observed both for recombinant proteins as well as during viral infection (5, 29). As with other coronaviruses, dimerization and oligomerization of the SARS-N protein has been extensively studied. Surjit et al. (38) investigated the self association of SARS nucleocapsid protein through a yeast two hybrid assay analysis. They mapped the self association site to the C-terminal 209 amino acids. In a similar study, Yu et al. (45) showed that recombinant N protein is capable of forming dimers and higher weight oligomers at high concentrations. The dimerization domain was determined to be at the C-terminal 138 residues (45). Another study conducted with He et al determined that the serine-arginine rich motif (SSRSSSRSRGNSR) is crucial for N-N oligomers formation and deletion of this motif leads to loss of N-N interaction. They believe that this self association may play an important role in ribonucleoprotein formation and subsequently in the viral assembly process (10).

Interaction of SARS- N protein with viral RNA

The coronavirus N-protein interacts with the viral RNA genome to form a helical nucleocapsid structure. Specific interactions between the N protein and coronavirus leader RNA have been defined. Anti-N antibodies precipitate full length as well as

subgenomic RNAs from infected cells (1) This interaction plays a role in a number of viral processes including viral RNA dependent RNA transcription, genome replication and translation (31). Synthesis of coronavirus mRNA takes place via a discontinuous transcription process that results in fusion of a leader RNA sequence to the subgenomic RNA body sequence. Although several models have been proposed for this procedure, the exact mechanism of action for this process is still unknown. Nucleic acid binding activity of Coronavirus N has been studied for MHV, BCV and IBV. This interaction was found to be specific to viral RNA and was not affected by the addition of non specific competitor RNA (36). Binding of the N protein to the viral RNAs occurs via the leader sequence since this is the common component of these RNAs. Studies with IBV N-protein have determined the presence of two RNA binding regions of the N protein; one in the C-terminal and the other at the N-terminal region of the N-protein that interact with the genomic RNA. However, RNA binding domain of MHV-N protein was determined to be within 136-397 residues (25). Nelson and Stohlman (30) further determined the binding domain to 169-308 amino acid residues of the MHV N-protein. Using chimeric BCV and MHV N-proteins, the RNA binding domain was determined to be between residues 163-380 (32). Using bacterial expression of MHV-N protein and RNA probes, the region of interaction for RNA and N protein binding was finally determined to be in a central 50 amino acid residues (177-231) of this fragment. This fragment is found to be rich in serine and arginine residues (31). This motif is conserved in all coronavirus N proteins, including SARS-CoV N protein. Using NMR, Huang et al. (13) recently determined the structure of the N-terminal binding domain of the SARS-N protein, specifically residues 45-181. This N-terminal domain had very little structural similarity

to any other coronavirus N protein. However, upon visualizing the three dimensional structure, it was determined that this domain had remarkable similarities to other ribonucleoprotein RNA binding proteins. These similarities included presence of positively charged lysine and arginine residues on the surface. Addition of untranslated viral 3′ RNAs confirmed the binding capacity of this domain. The authors also found various small molecules capable of binding to this region by NMR based screening (13). The interaction of the N protein is believed to take place due to the chemical exchange between the lysine and arginine residues and the phosphate backbone of the RNA molecules. Therefore, despite low homology with other coronavirus N-proteins, the RNA binding region of SARS-CoV N-protein consists of structural features consistent with RNA binding proteins (14).

Interaction of SARS-N with cellular proteins

The association of hnRNP-A1 with N-protein has been established in MHV nucleocapsid protein studies. Using mammalian expression systems, hnRNP-A1 and viral N-protein were expressed and the *in vitro* as well as *in vivo* association of these proteins was determined. It was determined, using yeast two hybrid assays that these proteins physically associated in a specific manner and independent of their association with the viral RNA. Using a series of deletion mutants, the binding site of hnRNP-A1 with MHV N-protein was determined in two regions of the protein using yeast two hybrid assays. The N-protein was associated with RNP-A1 via its N-terminal region (1-292) and at another site at the carboxyl terminal end (391-455). This association is believed to play

an important role in regulation of MHV RNA synthesis (41). Recently, the association between SARS-N protein and hnRNP-A1 has been studied using biophysical and biochemical assays (22). It was determined that SARS N-protein has a high binding affinity for hnRNP-A1. The association was mapped to the 161-210 amino acid fragment of SARS-N and to the 203-320 glycine rich region of hnRNP-A1. The authors postulated that this association plays an important role in SARS-CoV RNA synthesis.

Another cellular protein that is associated with SARS-N is human Cyclophilin A (hCypA). Cyclophilins were initially discovered because of their affinity to the immunosuppressive drug Cyclosporine A. More recently, it has been determined that cyclophilins play an important role in HIV infection via their association with the HIV-1 Gag polyprotein (37). Sequence comparison of SARS-N protein with the HIV capsid domain of the HIV-Gag protein shows moderate similarity (36.7%). The binding of Cyclophilin A with SARS N protein was confirmed using surface plasmon resonance (SPR) biosensor technology (22, 23). SARS-N protein associates with Cyclophilin A via the Trp-302-Pro310 loop, of which the Trp302 residue was major requirement recognition. The significance of this association is not established. However, HIV studies have reflected on decrease in viral entry and infection upon addition of CypA inhibitor CsA. Fig 1.3 describes a schematic representation of the various domains of SARS-N.

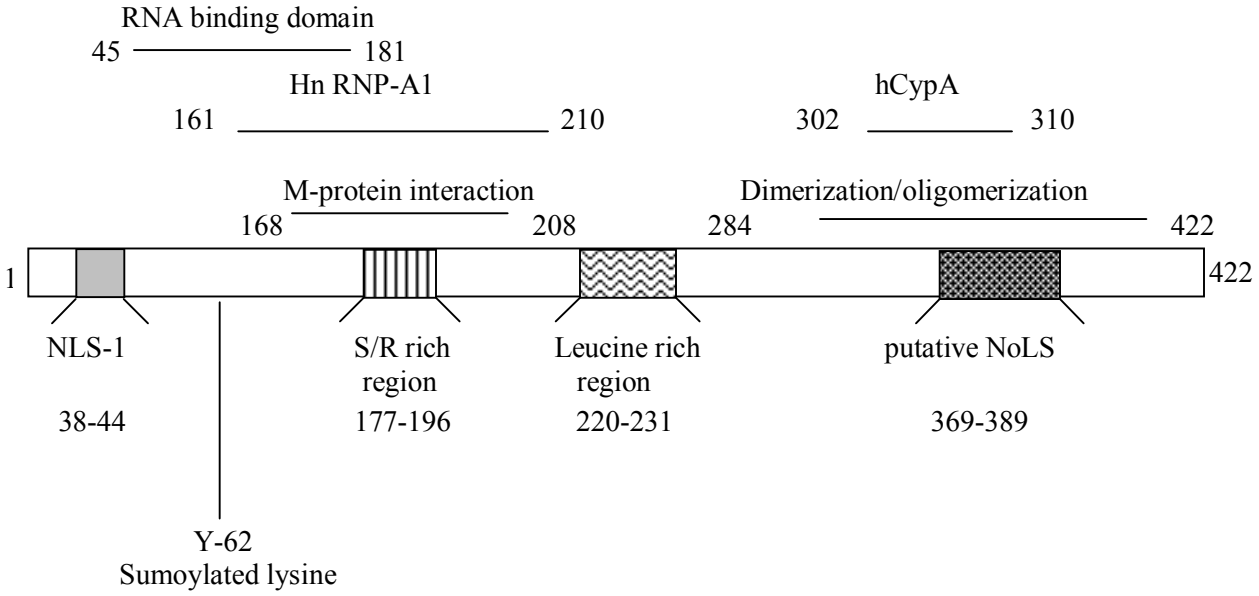


Fig 1.4 Diagrammatic representation of SARS- N domains.

Posttranslational modifications of SARS-N protein:

Several coronavirus N proteins have shown to be phosphorylated including IBV, MHV and TGEV N proteins even though the exact sites of phosphorylation of these have not been identified (3). Although the exact role of N protein phosphorylation is unknown, phosphorylation is postulated to play a role in determining its binding affinity to viral RNA. Using surface plasma resonance, it was demonstrated that phosphorylation of N protein decreased its binding affinity for random RNA. Phosphorylation may thus lead to the preferential binding of viral RNA over cellular RNA (3).

Dephosphorylation of coronavirus N protein takes place with the help of cellular phosphoprotein phosphatases and dephosphorylation facilitates viral infection by promoting dissociation of the N protein from the viral ribonucleoprotein complex. (26)

Serine phosphorylation of SARS-CoV N protein was initially reported by Surjit et al. (38). They determined that the phosphorylated SARS-N predominantly localized in the cytoplasm of the cells and associated with the membrane. They also demonstrated the phosphorylation dependent binding of N protein with 14-3-3 and postulate that this interaction facilitates the cytoplasmic localization. The inhibition of 14-3-3 expression by RNA interference leads to SARS-N accumulation in the nucleus of the cell (39).

Recently, Li et al. (18) reported sumoylation of the SARS-N protein. Because the N protein does not contain any cysteine residues, the dimerization and oligomerization of SARS-N must be due to other interactions. Sumoylation is described as the covalent attachment of a small ubiquitin like modifier (SUMO) to the protein. The sumoylation site of SARS-N was mapped to the lysine residue at position 62. Mutation of the K62 residue (K62A) using site directed mutagenesis demonstrated loss in SARS-N oligomerization and this led to the conclusion that sumoylation drastically promotes homo-oligomerization of the SARS-N protein. (18).

Intracellular localization properties of coronavirus N proteins

One of the foremost ways to determine the functional significance of a protein is to determine its localization properties inside the cells. The N protein is primarily found in the cytoplasm, which is consistent with role as a structural protein. Studies of the IBV N protein showed significant accumulation of N protein in the nucleus. Confocal microscopic analysis determined that this sub-nuclear structure was the nucleolus of the cell (12). Further studies determined that nucleolar localization was also a common feature among other members of the coronavirus family. Nuclear localization was observed in TGEV, a member of group I and MHV, a member of group II of the family. These proteins were found to co-localize with nucleolin, a nucleolar protein and was believed to abrogate cell division (43). Co-localization with fibrillarin was also observed (4).

Translocation of nuclear proteins from the cytoplasm, across the nuclear pore complex (NPC), and into the nucleus is initiated by the binding of a nuclear localization signal (NLS) on the cargo protein to importin-α. The cargo/importin-α complex , binds to a second shuttle protein, importin-β, which is then targeted to the NPC. In the presence of additional accessory proteins, including Ran-GDP, the cargo/importin complex is transported across the NPC and released into the nucleoplasm (8). Classical NLS sequences incorporate regions enriched in basic amino acids and generally conform to one of three types (11, 27). The "pat4" NLS consists of a continuous stretch of four basic amino acids (lysine or arginine) or three basic amino acids associated with histidine or

proline. The "pat7" NLS starts with a proline and is followed within three residues by a segment containing three basic residues out of four. The third type of NLS, known as a "bipartite" motif, consists of two basic amino acids, a 10 amino acid spacer, and a five amino acid segment containing at least three basic residues. Nucleolar localization results due to presence of the nucleolar localization signal (NoLS) in combination with classic NLSs and nuclear targeting signals (NTS). Nucleolar localization is usually preceded by nuclear translocation following targeting to the nucleolus.

The localization domains of SARS-CoV N protein were determined by PSORT computer program. For the purpose of comparison, we also analyzed N proteins of other coronaviruses. As described in Table 1, the available sequence information from GenBank indicates that all coronavirus N proteins possess at least one localization domain. Furthermore, these domains are generally confined to one of three regions within the protein; residues 14 to 60, 200-280, and 326-400. Within this group, SARS-CoV is the only N protein with NLS sequences in all three regions. The SARS-CoV N protein has two pat4 and four pat7 localization domains. Furthermore, besides FECV, SARS-CoV was the only known coronavirus to possess a bipartite signal sequence.

Table 1: Predicted NLSs by PSORT analysis

Virus Strain Accession number	Localization signals predicted by PSORT			
	GenBank No.	Pat 4	Pat 7	Bipartite
Coronavirus Gp I				
Transmissible gastroenteritis virus (TGEV)	YOO542	340-RKRK	NONE	NONE
HCV-Human Coronavirus (HCV-229E)	NP073556	NONE	45-PINKKDK 198-PQEKDKK 247-PRWKRQP	NONE
Feline coronavirus (FECV)	CAA74230	NONE	14-PSKRRGR	326-KRPSEVAKD QRQRKSRS
Coronavirus Gp II				
Human coronavirus (HCV-OC43)	NC 005147	NONE	270-PRQKRSP	NONE
Murine Hepatitis virus (MHV-JHM)	M25875	NONE	271-PRQKRTP 391-PKPQRKR 393-PRKRGT	NONE
Bovine coronavirus (BCoV-ENT)	NC_003045	NONE	270-PRQKRSP	NONE
Porcine Hemagglutinating encephalomyelitis virus (PHEV)	AAL80036	NONE	270-PRQKRSP	NONE
Severe acute respiratory syndrome coronavirus (SARS-CoV-Urbani)	AY278741	257-KKPR 373-KKKR	38-PKQRRPQ 259-PRQKRTA 369-PKKDKKK 384-PQRQKKQ	373-KKKKTDEAQ PLPQRQKK 374-KKKTDEAQP LPQRQKKQ
Coronavirus gp III				
IBV-Avian infectious Bronchitis virus-Beaudette	NC_001451	359-RPKK	360-PKKEKKL	NONE
Turkey coronavirus (TCV)	AAF23873	359-RPKK 364-KKPK 365-KPKK	360-PKKEKKP	NONE
Family *Arteriviridae* Porcine reproductive and respiratory syndrome virus (PRRSV-PA8)	AF176348	10-KRKK	41-PGKKNKK	NONE

Immunogenic properties of SARS-N protein

Coronavirus N protein is highly immunogenic and humoral response for SARS-N was observed for over 90% of patients infected with SARS-CoV (17, 20). Additionally, studies involving a plasmid encoding full-length SARS-N have been utilized to study its effect as a DNA vaccine (46). This vaccination in BALB/c mice led to strong antibody response involving increased cytokine secretion specifically that of IFN-γ and IL-2. More importantly, these vaccinated mice exhibited delayed hypersensitivity reactions and cytotoxic T cell response. A different study conducted by Liao et al (21) analyzed the interaction of SARS-CoV N protein and NF-κB. The transcription factor NF-κB activates several genes involved both in inflammation as well as cell mediated immune response including TNF- α, IL-2, IL-6 and IL-8. Liao et al. observed that full length N protein significantly increased NF-κB activity in a dose dependent manner. The humoral and cell mediated immune reactions in response to SARS-CoV N protein indicate that this protein may be involved in the viral pathogenesis of SARS. This protein can be used for development of vaccination and therapeutics for the treatment of SARS.

In conclusion, SARS-CoV N protein does not show significant similarities to any known coronavirus N proteins. Presence of unique serine-arginine rich region, multiple NLSs as well as unique cellular localization pattern make this protein an interesting study subject. Research involving molecular mechanisms of SARS-CoV N protein holds great promise for designing therapeutics as well as diagnostic tools for the virus infection.

References

1. **Baric R.S., G.W. Nelson, J.O. Fleming, R.J. Deans, J.G. Keck, N. Casteel and S.A. Stohlman. 1988.** Interaction between coronavirus nucleocapsid protein and viral RNAs: implications for viral transcription. J Virol. **62(11)**: 4280-87.
2. **Chang CK**, **Sue SC**, **Yu TH**, **Hsieh CM**, **Tsai CK**, **Chiang YC**, **Lee SJ**, **Hsiao HH**, **Wu WJ**, **Chang WL**, **Lin CH**, **Huang TH**. 2006. Modular organization of SARS coronavirus nucleocapsid protein. J Biomed Sci. Jan;13(1):59-72. Epub 2005 Oct 14.
3. **Chen H., A. Gill , B.K. Dove , S.R. Emmett, C.F. Kemp, M.A. Ritchie, M. Dee , J.A. Hiscox. 2005.** Mass spectroscopic characterization of the coronavirus infectious bronchitis virus nucleoprotein and elucidation of the role of phosphorylation in RNA binding by using surface plasmon resonance**.** J Virol. **79(2)**:1164-79.
4. **Chen H., T. Wurm , P. Britton, G. Brooks, J.A. Hiscox. 2002.** Interaction of the coronavirus nucleoprotein with nucleolar antigens and the host cell. J Virol. **76(10)**:5233-50.
5. **Deregt D., M. Sabara , L.A. Babiuk . 1987.** Structural proteins of bovine coronavirus and their intracellular processing. J Gen Virol. **68 (Pt 11)**:2863-77.
6. **Duan S.M., X.S. Zhao , R.F. Wen, J.J Huang, G.H. Pi , S.X. Zhang, J. Han, S.L. Bi , L. Ruan, X.P. Dong ; SARS Research Team. 2003.** Stability of SARS coronavirus in human specimens and environment and its sensitivity to heating and UV irradiation. Biomed Environ Sci. **16(3)**:246-55.

7. **Escors D., J. Ortego, H. Laude, L. Enjuanes. 2001.** The membrane M protein carboxy terminus binds to transmissible gastroenteritis coronavirus core and contributes to core stability. J Virol. **75(3)**:1312-24.
8. **Gorlich, D., and U. Kutay. 1999.** Transport between the cell nucleus and the cytoplasm. Ann. Rev. Cell. Dev. Biol. **5**:607-660.
9. **Guan Y., B.J. Zheng, Y.Q. He**, **X.L. Liu** , **Z.X. Zhuang**, **C.L. Cheung**, **S.W. Luo**, **P.H. Li** , **L.J. Zhang** , **Y.J. Guan** , **K.M Butt** , **K.L. Wong** , **K.W. Chan** , **W. Lim** , **K.F Shortridge** , **K.Y. Yuen** , **J.S. Peiris**, **L.L Poon**. **2003**. Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. Science. **10; 302(5643)**:276-8.
10. **He R., A. Leeson, M. Ballantine , A. Andonov, L. Baker, F. Dobie, Y. Li , N. Bastien, H. Feldmann , U. Strocher, S. Theriault, T. Cutts, J. Cao, T.F. Booth, F.A. Plummer, S. Tyler, X. Li . 2004.** Characterization of protein-protein interactions between the nucleocapsid protein and membrane protein of the SARS coronavirus. Virus Res. **105(2)**:121-5.
11. **Hicks, G. R., and N. V. Raikhel. 1995.** Protein import into the nucleus: an integrated view, Ann. Rev. Cell Dev. Biol. **11**:155-188.
12. **Hiscox J.A., T. Wurm , L. Wilson , P. Britton , D. Cavanagh , G. Brooks . 2001.** The coronavirus infectious bronchitis virus nucleoprotein localizes to the nucleolus. J.Virol. **75(1)**:506-12.
13. **Huang Q.**, **L. Yu**, **A.M. Petros**, **A. Gunasekera**, **Z. Liu** , **N. Xu**, **P. Hajduk**, **J. Mack**, **S.W. Fesik**, **E.T. Olejniczak**. **2004a.** Structure of the N-terminal RNA-

binding domain of the SARS CoV nucleocapsid protein. Biochemistry. **43(20)**:6059-63.

14. **Huang Y., Z.Y.Yang, W.P. Kong , G.J. Nabel. 2004b.** Generation of synthetic severe acute respiratory syndrome coronavirus pseudoparticles: implications for assembly and vaccine production. J Virol. **78(22)**:12557-65.
15. **Kuo L., P.S. Masters. 2002.** Genetic evidence for a structural interaction between the carboxy termini of the membrane and nucleocapsid proteins of mouse hepatitis virus. J Virol. **76(10)**:4987-99.
16. **Lau S.K., P.C. Woo , K.S. Li, Y. Huang, H.W. Tsoi, B.H. Wong, S.S. Wong, S.Y. Leung, K.H. Chan, K.Y. Yuen. 2005.** Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. Proc Natl Acad Sci U S A. **102(39)**:14040-5.
17. **Leung D.T., F.C. Tam , C.H. Ma, P.K. Chan, J.L. Cheung, H. Niu, J.S. Tam , P.L. Lim . 2004.** Antibody response of patients with severe acute respiratory syndrome (SARS) targets the viral nucleocapsid. J Infect Dis. **190(2)**:379-86.
18. **Li F.Q., H. Xiao, J.P. Tam, D.X. Liu. 2005a.** Sumoylation of the nucleocapsid protein of severe acute respiratory syndrome coronavirus. FEBS Lett. **579(11)**:2387-96.
19. **Li W., Z. Shi, M. Yu , W. Ren, C. Smith, J.H. Epstein, H. Wang, G. Crameri , Z. Hu, H. Zhang, J. Zhang, J. McEachern, H. Field , P. Daszak , B.T. Eaton, S. Zhang, L.F. Wang. 2005b.** Bats are natural reservoirs of SARS-like coronaviruses. Science. **310(5748)**:676-9. Epub 2005 Sep 29.

20. **Li Y. H., J. Li, X.E. Liu, L. Wang, T. Li, Y.H. Zhou, H. Zhuang. 2005c.** Detection of the nucleocapsid protein of severe acute respiratory syndrome coronavirus in serum: Comparison with results of other viral markers. J Virol Methods. **130(1-2)**:45-50**.**

21. **Liao Q., L. Ye, K.A. Timani, Y.C. Zeng, Y.L. She, Y. Li and Z. Wu. 2005.** Activation of NF-κB by the full length Nucleocapsid protein of the SARS coronavirus. Acta Biochimica et Biophysica Sinica. **37(9)**: 607-612

22. **Luo C., H. Luo and S. Zheng. 2004.** Nucleocapsid protein of SARS coronavirus tightly binds to human cyclophilin A. Biochem Biophys Res Commun. **321(3)**:557-65.

23. **Luo H., Q. Chen, J. Chen , K. Chen, X. Shen and H. Jiang 2005.** The nucleocapsid protein of SARS coronavirus has a high binding affinity to the human cellular heterogeneous nuclear ribonucleoprotein A1. FEBS Lett. **579(12)**:2623-8.

24. **Marra M.A., S.J. Jones and C.R. Astell et al. 2003.** The Genome sequence of the SARS-associated coronavirus. Science. **300(5624)**:1399-404.

25. **Masters P.S. 1992.** Localization of an RNA-binding domain in the nucleocapsid protein of the coronavirus mouse hepatitis virus. Arch Virol. **125(1-4)**:141-60.

26. **Mohandas D.V. and S. Dales . 1991.** Endosomal association of a protein phosphatase with high dephosphorylating activity against a coronavirus nucleocapsid protein. FEBS Lett. **282(2)**:419-24.

27. **Nakai, K.and M. Kanehisa. 1992.** A knowledge base for predicting protein localization sites in eukaryotic cells. Genomics **14**:897-911.

28. **Narayanan K., A. Maeda, J. Maeda and S. Makino. 2000.** Characterization of the coronavirus M protein and nucleocapsid interaction in infected cells. J Virol. **74(17)**:8127-34.

29. **Narayanan K., K.H. Kim and S. Makino. 2003.** Characterization of N protein self-association in coronavirus ribonucleoprotein complexes. Virus Res. **98(2)**:131-40.

30. **Nelson G.W. and S.A. Stohlman . 1993.** Localization of the RNA-binding domain of mouse hepatitis virus nucleocapsid protein. J Gen Virol. **74 (Pt 9)**:1975-9.

31. **Nelson G.W., S.A. Stohlman and S.M. Tahara. 2000.** High affinity interaction between nucleocapsid protein and leader/intergenic sequence of mouse hepatitis virus RNA. J Gen Virol. **81**: 181-8.

32. **Peng D., C.A. Koetzner and P.S. Masters. 1995.** Analysis of second-site revertants of a murine coronavirus nucleocapsid protein deletion mutant and construction of nucleocapsid protein mutants by targeted RNA recombination. J Virol. **69(6)**:3449-57.

33. **Rota P.A., M.S. Oberste and Monroe S.S. 2003.** Characterization of a novel coronavirus associated with severe acute respiratory syndrome. Science. **300 (5624)**: 1377-8.

34. **Schaad M.C., W. Chen, S.A. Peel and R.S. Baric. 1993.** Studies into the mechanism for MHV transcription. Adv Exp Med Biol. **342**:85-90.

35. **Snijder E.J., P.J. Bredenbeek , J.C. Dobbe, V. Thiel, J. Ziebuhr, L.L. Poon , Y. Guan, M. Rozanov, W.J. Spaan and A.E. Gorbalenya. 2003.** Unique and

conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. J Mol Biol. **331(5)**:991-1004.

36. **Stohlman S.A., R.S. Baric, G.N. Nelson, L.H. Soe, L.M. Welter and R.J. Deans. 1988.** Specific interaction between coronavirus leader RNA and nucleocapsid protein. J Virol. **62(11)**:4288-95.
37. **Streblow D.N., M. Kitabwalla and C.D. Pauza. 1996.** Gag protein from human immunodeficiency virus type 1 assembles in the absence of cyclophilin A. Virology. **252(1)**:228-34.
38. **Surjit M., B. Liu , P. Kumar, V.T. Chow and S.K. Lal. 2004.** The nucleocapsid protein of the SARS coronavirus is capable of self-association through a C-terminal 209 amino acid interaction domain. Biochem Biophys Res Commun. **317(4)**:1030-6.
39. **Surjit M., R. Kumar, R.N. Mishra, M.K. Reddy, V.T. Chow and S.K. Lal . 2005.** The severe acute respiratory syndrome coronavirus nucleocapsid protein is phosphorylated and localizes in the cytoplasm by 14-3-3-mediated translocation. J Virol. **79(17)**:11476-86.
40. **Tang T.K., M.P. Wu , S.T. Chen, M.H. Hou, M.H. Hong , F.M. Pan, H.M. Yu , J.H. Chen, C.W. Yao and A.H. Wang. 2005.** Biochemical and immunological studies of nucleocapsid proteins of severe acute respiratory syndrome and 229E human coronaviruses. Proteomics. **5(4)**:925-37.
41. **Wang Y. and X. Zhang. 1999.** The nucleocapsid protein of coronavirus mouse hepatitis virus interacts with the cellular heterogeneous nuclear ribonucleoprotein A1 in vitro and in vivo. Virology. **265(1)**:96-109.

42. **Wang Y., X. Wu, Y. Wang, B. Li, H. Zhou, G. Yuan, Y. Fu and Y. Luo. 2004.** Low stability of nucleocapsid protein in SARS virus. Biochemistry. **43(34)**:11103-8.

43. **Wurm T., H. Chen, T. Hodgson, P. Britton, G. Brooks and J.A. Hiscox. 2001.** Localization to the nucleolus is a common feature of coronavirus nucleoproteins, and the protein may disrupt host cell division. J Virol. **75(19)**:9345-56.

44. **Ying W. et al. 2004.** Proteomic analysis on structural proteins of Severe Acute Respiratory Syndrome coronavirus. Proteomics. **4(2)**:492-504.

45. **Yu I.M., C.L. Gustafson, J. Diao, J.W. Burgner 2nd, Z. Li, J. Zhang and J. Chen . 2005.** Recombinant severe acute respiratory syndrome (SARS) coronavirus nucleocapsid protein forms a dimer through its C-terminal domain. J Biol Chem. **280(24)**:23280-6.

46. **Zhao P., J. Cao, L.J. Zhao, Z.L. Qin, J.S. Ke, W. Pan, H. Ren, J.G. Yu and Z. T. Qi. 2005.** Immune responses against SARS-coronavirus nucleocapsid protein induced by DNA vaccine. Virology. **331(1)**:128-35.

Section 2

Intracellular localization of the SARS coronavirus nucleocapsid protein: absence of nucleolar accumulation during infection and after expression as a recombinant protein in Vero cells.

ABSTRACT

The nucleocapsid (N) protein of several members within the order *Nidovirales* localizes to the nucleolus during infection and after transfection of cells with N genes. However, confocal microscopy of N protein localization of Vero cells infected with the severe acute respiratory syndrome coronavirus (SARS-CoV) or transfected with the SARS-CoV N gene failed to show the presence of N in the nucleoplasm or nucleolus. Amino acids 369-389, which contain putative nuclear localization signal (NLS) and nucleolar localization signal (NoLS) motifs, failed to restore nuclear localization to an NLS-minus mutant Rev protein. These data indicate that nuclear localization is not a conserved property within the *Nidovirales*.

Introduction

Severe acute respiratory syndrome (SARS) is caused by an enveloped, positive-stranded RNA virus, the SARS coronavirus (SARS-CoV). Along with the arteriviruses, toroviruses and roniviruses , the coronaviruses are placed in a single order, *Nidovirales* (Cavanagh, 1997). Even though nidovirus replication is restricted to the cytoplasm, the nucleocapsid protein (N) of two arteriviruses, porcine reproductive and respiratory syndrome virus (PRRSV) and equine arteritis virus (EAV), have been reported to localize to the nucleolus during infection (19,20,23,25). An identical cytoplasmic-nucleolar distribution pattern has also been reported for the N proteins of several coronaviruses, including representative members of group I (transmissible gastroenteritis virus; TGEV), group II (mouse hepatitis virus; MHV), and group III (infectious bronchitis virus; IBV) viruses. These arterivirus and coronavirus N proteins, when expressed alone or fused to the red-shifted enhanced green fluorescent protein (EGFP), also localize to the nucleolus, demonstrating that translocation across the nuclear pore complex (NPC) and accumulation in the nucleolus are independent of other viral proteins (11, 20, 23). Once inside the nucleolus, the PRRSV and IBV N proteins co-localize with major nucleolar proteins, including nucleolin and fibrillarin (4, 25). The capacity of N to modulate nucleolar function may represent a viral strategy that diverts biosynthetic resources from the dividing nucleus to the cytoplasm, the site of virus replication (10, 11, 21, 25). There also appears to be a mechanism for the export of N from the nucleolus and back to the cytoplasm. The N proteins of PRRSV and EAV accumulate in the nucleus in response to

the treatment of cells with lemptomycin B (LMB; 21, 23), an inhibitor of the export shuttle protein, CRM1. (7).

The classical scheme for the translocation of a protein from the cytoplasm through the nuclear pore complex and into the nucleoplasm is initiated through the interaction between a nuclear localization signal (NLS) on the cargo protein with the NLS-binding site on importin-a (reviewed in 2,8). Classical NLS sequences are enriched in basic amino acids, such as lysines and arginines, and generally conform to one of three types, known as monopartite (pat4and pat7) and bipartite motifs (9,17) The pat4 NLS is defined as a continuous stretch of four basic amino acids (lysine or arginine) or three basic amino acids associated with histidine or proline. The pat7 NLS starts with proline followed within three residues by a segment containing three basic residues out of four. The "bipartite" motif, consists of two basic amino acids, a ten amino acid spacer, and a five amino acid segment containing at least three basic residues (reviewed in 9 and 17). The nuclear transport of the PRRSV N protein is dependent on a single pat7 NLS, 41-PGKKNKK, which overlaps the RNA binding domain (20, 21, 22, 25). Site-directed scanning mutagenesis identified the requirement of the four lysine residues, 43-KK and 46-KK, for the transport of N into the nucleus (21). A second NLS, 10-KRKK, is located upstream and functions as a cryptic NLS, which becomes accessible following a conformational change in the N protein (20, 21). The nucleolus is not a membrane-bound organelle; therefore, the localization of a protein to the nucleolus occurs by diffusion through the nucleoplasm and accumulation in the nucleolus via a trans-acting nucleolar targeting signal (NoTS). Viral proteins frequently combine an NLS and NoTS into a

single localization signal sequence, often referred to as a compact nucleolar localization signal (NoLS) sequence (14, 15). Compact NoLS sequences are typically no longer than 30 amino acids and possess at least nine basic amino acids, including at least one NLS motif (14).

Results and Discussion

When analyzed for the presence of NLS-like signal sequences using the computer program PSORT (17), the 422 amino acid SARS-CoV N protein is shown to possess as many as eight NLS motifs, which are distributed between amino acids 38-44, 257-265 and 369-389 (Fig. 2.1). Embedded within the 369-389 domain are five NLS motifs, including a lysine enriched peptide sequence, 369-PKKDKKK-375, which possesses similarities to the pat7 NLS in PRRSV N (20) and the well characterized NLS, 126-PKKKRLV, found in the SV40 large T antigen (13). In addition, the 369-375 region resembles the putative NLS sequence, 360-PKKEKKL, reported for the IBV N protein (24). The 369-389 region shares similarities with NoLS domains found in nucleolar proteins of other viral proteins, including HIV Rev, HIV Tat, HTLV Rex, and the capsid protein of Semliki Forest virus (6, 14, 15, 20).

Fig. 2.1. NLS motifs in SARS-CoV N protein. Peptide sequence analysis of the SARS-CoV N protein Urbani isolate (GenBank accession no. AY278741) using the Web-based program, PSORT (17). The rules for defining pat4, pat7 and bipartite motifs are discussed in the introductory comments of the text.

38-PKQRRPQ (pat7)

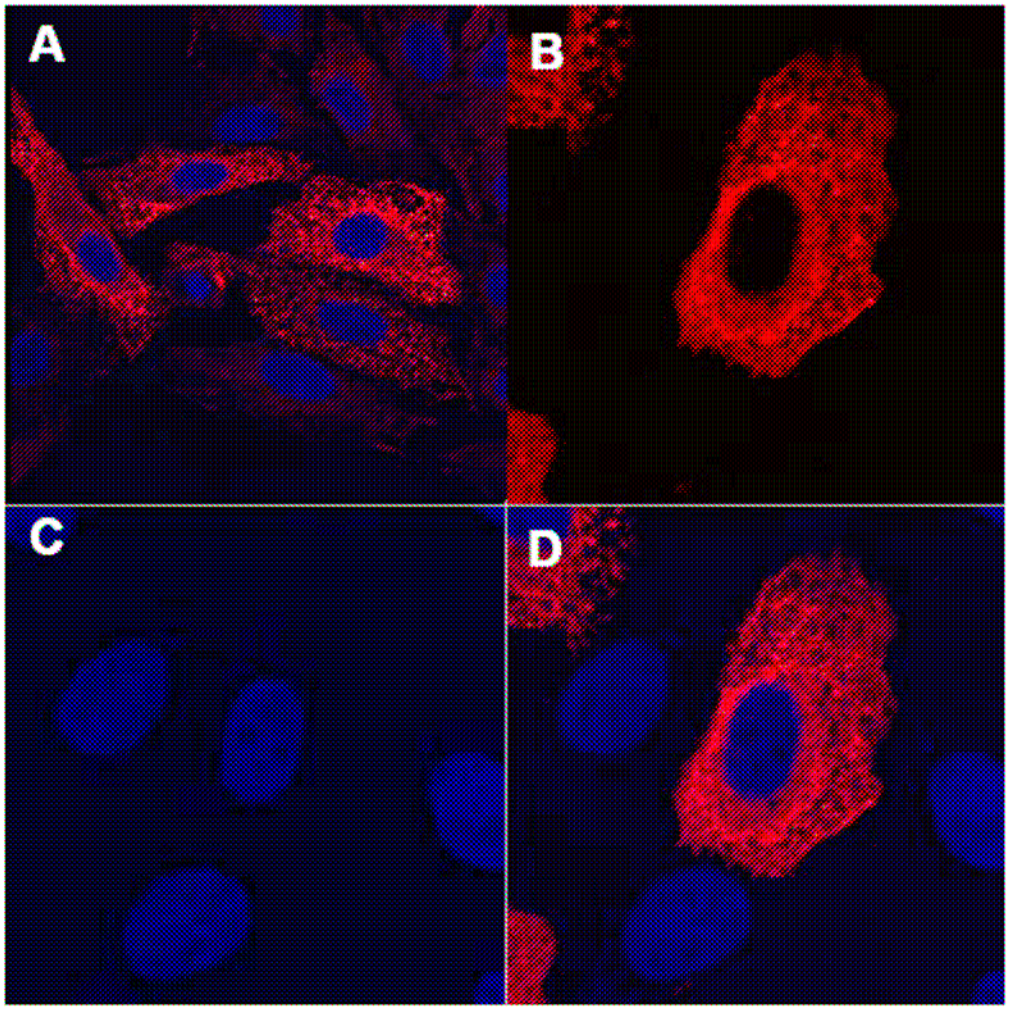
257-KKPRQKRTAKKPR (pat4)
PRQKRTA (pat7)

369-PKKDKKKKTDEAQPLPQRQKKQKKKK (pat4)
PKKDKKKK (pat7)
PQRQKKQ (pat7)
KKKKTDEAQPLPQRQKK (bipartite)
KKKTDEAQPLPQRQKKQ (bipartite)

The localization of SARS-CoV N during infection was studied using confocal microscopy of infected cells stained with SA 46-4, a SARS-CoV N protein-specific mAb prepared in our laboratory against recombinant N prepared from the SARS-CoV Urbani isolate. Nuclei were identified using the nucleic acid stain, TO-PRO-3 (Molecular Probes; Maziere et al., 1996). Vero cells (ATCC), grown on coverslips, were mock-infected or infected with SARS-CoV (Urbani isolate) at a multiplicity of infection of 0.1. Coverslips were removed at 6, 12 and 24 h after infection, washed with PBS, fixed for 10 minutes with 2% paraformaldehyde in PBS, and then stained with SA 46-4 followed by anti-mouse IgG antibody conjugated to AlexaFluor 594 (Molecular Probes). All antibodies were diluted 1:500 in blocking buffer (PBS containing 3% normal goat sera and 0.2% saponin) and incubations performed for 1 h at room temperature followed by extensive washing in PBS. Cells were counterstained with TO-PRO-3, diluted in blocking buffer, then viewed on a Zeiss LSM 510 confocal microscope. In each experiment at least 50

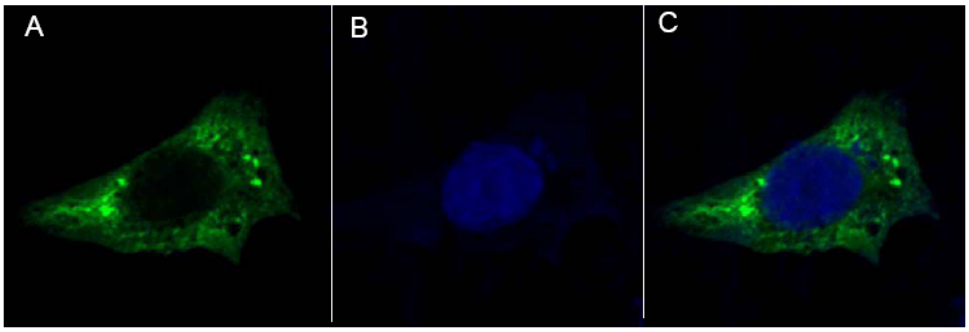
infected cells in 10 distinct fields were analyzed. Representative results at 24 h after infection are presented in Fig. 2.2. A group of infected cells, presented in Fig. 2.2A, shows no evidence for anti-N staining in the nucleus and nucleolus. A higher magnification of a representative cell is presented in Fig. 2.2 B, C, D. We found no co-localization between SA 46-4 staining and the nuclear stain, TO-PRO-3. The same results were obtained from cells fixed and stained with SA 46-4 and TOPRO-3 at 6 and 12 h after infection (data not shown). Wurm et al. (24) reported the nucleolar localization of coronavirus N proteins in Vero cells transfected with N genes from TGEV, MHV and IBV, and stained with anti-N antibodies. IBV N tagged with EGFP also localized to the nucleolus of Vero cells (11).

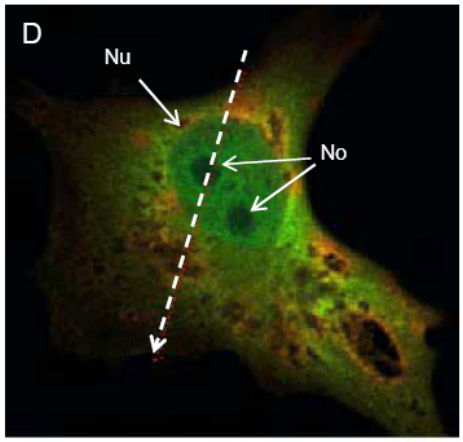
Fig.2.2. Distribution of N in SARS-CoV-infected cells. (A) Confocal microscopy of a cluster of infected cells at 24 h after infection. Viral antigen was detected using AlexaFluor 594-labeled SA 46-4 anti-N mAb (red). Cells were counterstained with the nuclear stain, TO-PRO-3 (blue). (B, C and D) are images of a single representative infected cell at 24 h after infection. (B) AlexaFluor 594-labeled SA 46-4. (C) TO-PRO-3 staining. (D) Merged image combining SA 46-4 and TO290 PRO-3 staining. An 0.8 um slice through the nucleus is shown in each image. Images were obtained with a 63x oil objective and panels B,C and D are magnified 2x

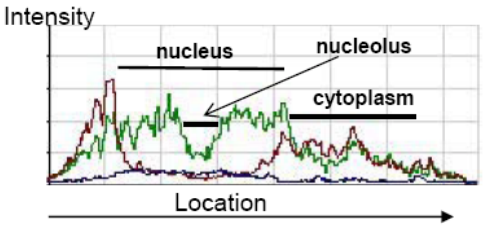


We performed similar studies of the SARS-CoV N gene expressed in Vero cells. A cDNA containing the N gene of SARS-CoV Urbani isolate was cloned into pIRES-EGFP, pEGFP-N1 and pEGFP-C2 eukaryotic expression vectors (Clontech). The pIRES-EGFP vector construct, pSARS-N-IRES-EGFP, through the incorporation of an internal ribosome entry site (IRES) separating the SARS-CoV N and EGFP genes, expresses EGFP and N as separate proteins. The other vectors express N as an EGFP fusion protein with the N protein fused in frame to the amino (SARS-N-EGFP) or carboxyl (EGFP-SARS-N) end of EGFP. The transfection of plasmid DNA into Vero cells was performed using Lipofectamine (Invitrogen) according to the manufacturer's directions. A representative cell at 24 hr after transfection pSARS-N-EGFP is shown in Fig. 2.3 panels A, B and C. Confocal microscopy showed that SARS-N-EGFP fluorescence was restricted to the cytoplasm with no visual evidence for N in the nucleolus. Similar results were obtained for Vero cells transfected with pSARS-N-IRES-EGFP and stained with the anti-N mAb, SA 46-4 (data not shown). Hiscox et al. (11) reported only a small percentage of cells with IBV N protein in the nucleus/nucleolus. In cells transfected with pSARS-N-EGFP or pEGFP-SARS-N, we noticed that a small number of cells contained at least some EGFP fluorescence in the nucleoplasm, but not in the nucleoli. A single cell exhibiting EGFP fluorescence in the nucleus and stained with SA 45-4 is shown in Fig. 2.3D. The intensity of red (SA 46-4) and green (EGFP) fluorescence taken along a single axis through the cytoplasm and nucleus is presented in the profile below the photomicrograph. The fluorescence intensity profiles show that red fluorescence is primarily restricted to the cytoplasm (see Fig. 2.3D); whereas, EGFP fluorescence can be found in both nuclear and cytoplasmic compartments, but is restricted from the nucleoli.

Fig 2.3 Localization of SARS-CoV N-EGFP in Vero cells. (A, B, C) Confocal microscopic images of a single cell at 24 hr after transfection with pSARS-N-EGFP. (A) EGFP fluorescence, (B) TO-PRO-3 fluorescence, (C) Merged image. (D) Image of a single cell transfected with pSARS-N-EGFP and exhibiting EGFP fluorescence in the nucleus. The cell was counterstained with AlexaFluor 594-labeled SA46-4 anti-N mAb (red). Below the image is a profile showing the intensity of EGFP (green line) and AlexaFluor 594 (red line) fluorescence along the axis shown by the dotted arrow.





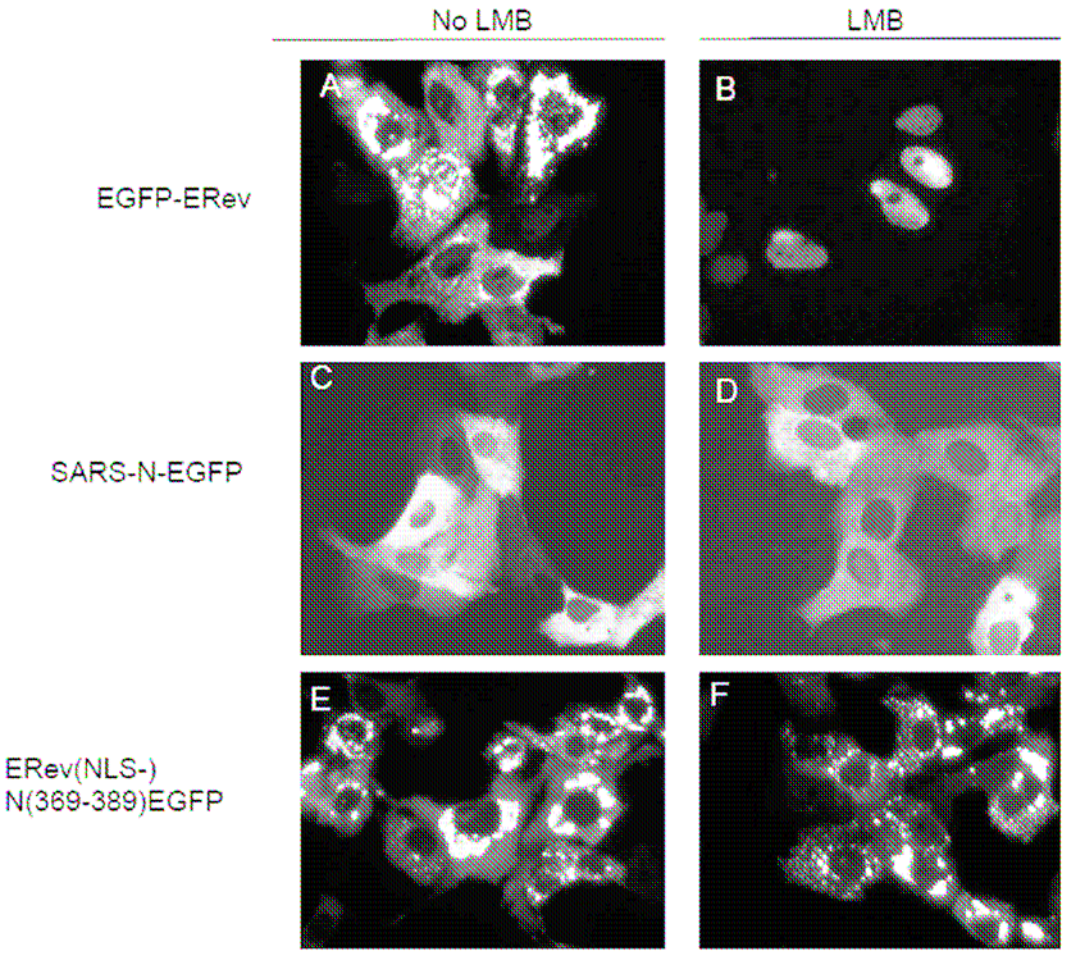


The possibility that EGFP fluorescence in the nucleus was the result of free EGFP was confirmed by performing a Western blot of Vero cells transfected with pSARS-N-EGFP or pEGFP-SARS-N constructs. Proteins in cell lysates were separated by SDS PAGE, transferred to a nylon membrane and probed with anti-GFP antibody. The blot showed the presence of full-length N-EGFP and EGFP-N fusion proteins, as well as smaller-sized products, including immunoreactive proteins which migrated the same as free EGFP (data not shown). Therefore, the presence of EGFP in the nucleus of some transfected cells is the likely result of the proteolytic cleavage of the SARS-N EGFP fusion protein and release of smaller sized EGFP-labeled fragments. Collectively, the results from infected and transfected cells indicate that SARS-CoV N does not appear to target the nucleolus or nucleus. One possible explanation for the absence of detectable amounts of N in the nucleus is that N enters the nucleus, but is rapidly exported. Therefore, under steady state conditions only a small quantity of N would be visible in the nucleus at any one time. In order to test this possibility, we treated pSARS-N-EGFP transfected cells with the nuclear export blocker, LMB, an antibiotic which inhibits CRM1, the major export shuttle protein (7, 18). LMB, at concentrations up to 50nM, was added to Vero cells between 4 and 18 h after transfection. After addition of LMB, the live cells were then followed under a fluorescence microscope for an additional 24 h. Cells transfected with a recombinant equine infectious anemia virus Rev (ERev) cDNA fused to EGFP(pEGFP-ERev) were included as a positive control. The continuous shuttling of ERev back and forth between the cytoplasm and the nucleus is dependent on single NLS and nuclear export signal (NES) domains (16, 18). The majority of Vero cells transfected with pERev-EGFP showed the presence of ERev-EGFP fusion protein in the cytoplasm

(Fig. 2.4A). In the presence of 10 nm LMB, the distribution pattern shifted towards the accumulation of ERev-EGFP in the nucleus (Fig. 2.4B). Accumulation of ERev-EGFP in the nucleus appeared as early as 6 h after the addition of LMB. Fig. 2.4C and D show representative results for cells transfected with pSARS-N EGFP and treated with LMB. The addition of LMB did not alter the cytoplasmic distribution of SARS-N-EGFP. The same result was obtained for cells transfected with pSARS-N-IRES-EGFP and stained with SA 46-4 (data not shown). LMB, when added as early as 4 h after transfection and at concentrations as high as 50 nM, did not affect the localization properties of SARS-N EGFP.

Fig. 2.4. Effect of LMB on the localization of SARS-N-EGFP in live cells. (A, B) Cells transfected with the control construct, pEGFP-ERev. (C, D) Cells transfected with pSARS-N EGFP. (E, F) Cells transfected with the Rev-SARS chimeric construct, pRev(NLS-minus)SARS N(368-389). LMB, at a concentration of 10 nM, was added at 4h after transfection. The results following LMB treatment are shown in the panels on the left. Photomicrographs were taken at 18 h after transfection.

Fig. 2.4. Effect of LMB on the localization of SARS-N-EGFP in live cells.



These data further support the notion that the SARS-CoV N protein is not translocated into the nucleus. Putative NLS sequences in the N proteins of IBV, TGE and MHV coronaviruses are found at amino acids 340-RKRK, 391-PKPQRKR and 360 PKKEKKL, respectively (11, 24). The corresponding region in SARS-CoV N protein is located in a lysine-rich region, between amino acids 369 and 389 (Fig. 2.1). Furthermore, the peptide sequence, 369-PKKDKKK of SARS-CoV N, is nearly identical to the well characterized NLS in PRRSV N protein (20) as well as a putative NLS in the N protein of IBV (24). To determine if the 369-389 peptide possessed nuclear transport activity, we tested the ability of the 369-389 peptide to substitute for the NLS of ERev. The pat4 NLS of ERev was removed by deleting six amino acids, 159-KRRRKHL, from the C-terminus. The construct, pERev(NLS-minus)N(369-389)EGFP, was made by ligating the cDNA corresponding to the SARS N 369-389 peptide to the C-terminal end of the mutant ERev and followed by EGFP. In Vero cells transfected with pERev(NLS-minus)N(369-389)EGFP, the chimeric EGFP protein localized to the cytoplasm (Fig. 2.4E). Blocking CRM1 by the addition of 10 nm LMB did not alter the cytoplasmic distribution of the chimeric protein (Fig. 2.4F). These data indicate that the 369-389 region of SARS-CoV N lacks the ability to substitute for the NLS of a known nucleocytoplasmic protein. Presumably, the translocation of TGE, MHV, and IBV coronavirus N proteins across the NPC and into the nucleoplasm is dependent on the activities of classical NLS domains (11, 24). The presence of eight NLS motifs, scattered between three different regions of the SARS-CoV N polypeptide provided a compelling argument for the nuclear and nucleolar localization of SARS CoV N. However, the results from this study of N protein localization in Vero cells infected with SARS-CoV or transfected with constructs

expressing the N gene showed no evidence for the localization of SARS-CoV N to the nucleoplasm or nucleolus. The absence of NLS activity within the lysine-rich 369-389 domain is puzzling. We can propose three mechanisms for the absence of NLS activity by the 369-389 peptide. First, even though NLS sequences frequently possess negatively charged residues, additional negative charges conferred by acidic amino acid residues or phosphorylated serine, threonine or tyrosine residues, can lower the affinity of the interaction between the NLS and NLS-binding site on importin-a (12). An aspartic acid residue is found at position 371, within the pat7 NLS motif of the SARS CoV-N. Additional, negatively charged amino acids are found at positions 378 and 379 (Fig.2.1). One more negative charge is made available through the potential phosphorylation of threonine at position 377 (1). Therefore, four negatively charged residues may be sufficient to neutralize both monopartitie and bipartitie NLS sequences within the 369-389 region. A second possibility is that the NLS is in a poor conformation to be recognized by importin-a. And finally, NLS recognition may be sterically blocked by the interaction between the 369-389 region and a cytoplasmic protein. In this study we did not evaluate the nuclear localization properties of the NLS motifs located in the other regions of the SARS-CoV N protein. Since the SARS-CoV N protein does not appear to enter the nucleus, we can assume that these NLS motifs are inactive, inaccessible or there are additional domains, which are responsible for retaining N in the cytoplasm.

References

1. Blom, N., S. Gammeltoft, and S. Brunak. 1999. Sequence- and structure-based prediction of eukaryotic protein phosphorylation sites. J. Mol. Biol. **294**:1351-1362.

2. **Catimel, B., T. Teh, M. R. Fontes, I. G. Jennings, D. A. Jans, G.J. Howlett, E.C. Nice and B. Kobe. 2001.** Biophysical characterization of interactions involving importin-alpha during nuclear import. J. Biol. Chem. **276**:34189-34198.

3. Cavanagh, D.1997. Nidovirales: a new order comprising Coronaviridae and Arteriviridae, Arch. Virol. **142**:629-633.

4. Chen, H., T. Wurm, P. Britton, G. Brooks, and J. A. Hiscox. 2002. May, Interaction of the coronavirus nucleoprotein with nucleolar antigens and the host cell, J. Virol. **76**:5233-5250.

5. De Maziere, A. M., W.J. Hage, and G. A Ubbels. 1996. A method for staining of cell nuclei in Xenopus laevis embryos with cyanine dyes for whole-mount confocal laser scanning microscopy. J. Histochem. Cytochem. **44**:399-402.

6. Favre, D., E. Studer, and M. R. Michel. 1994. Two nucleolar targeting signals present in the N-terminal part of Semliki Forest virus capsid protein. Arch. Virol. **137**:149-155.

7. Fornerod, M., M. Ohno, M. Yoshida, and I. Mattaj. 1997. CRM1 is an export receptor for leucine-rich nuclear export signals. Cell **90**:1051-1060.

8. Gorlich, D., and U. Kutay. 1999. Transport between the cell nucleus and the cytoplasm. Ann. Rev. Cell. Dev. Biol. **5**:607-660.

9. Hicks, G. R., and N. V. Raikhel. 1995. Protein import into the nucleus: an integrated view, Ann. Rev. Cell Dev. Biol. **11**:155-188.

10. Hiscox, J. A. 2002. The nucleolus--a gateway to viral infection? Arch. Virol. **47**:1077- 1089.

11. Hiscox, J. A., T. Wurm, L. Wilson, P. Britton, D. Cavanagh, and G. Brooks. 2001. The coronavirus infectious bronchitis virus nucleoprotein localizes to the nucleolus. J. Virol.**75**:506-512.

12. Jans, D.A., and S. Hubner. 1996. Regulation of protein transport to the nucleus: central role of phosphorylation. Physiol. Rev.**76**:651-685.

13. Kalderon, D., B. L. Roberts, W. D. Richardson, and A. E. Smith. 1984. A short amino acid sequence able to specify. Cell **39**: 499-509.

14. Kubota, S., T. D. Copeland, and R. J. Pomerantz. 1999. Nuclear and nucleolar targeting of human ribosomal protein S25: common features shared with HIV-1 regulatory proteins.Oncogenes **18**:1503-1514.

15. Kubota, S., H. Siomi, T. Satoh, S. Endo, M. Maki, and M. Hatanaka. 1989. Functional similarity of HIV-I rev and HTLV-I rex proteins: Identification of a new nucleolar targeting signal in rev protein. Biochem. Biophys. Res. Comm. **162**:963-970.

16. Meyer, B. E., J. L. Meinkoth and M. H. Malim. 1996. Nuclear transport of human immunodeficiency virus type 1, visna virus, and equine infectious anemia virus Rev proteins: identification of a family of transferable nuclear export signals. J. Virol. **70**:2350-2359.

17. Nakai, K.and M. Kanehisa. 1992. A knowledge base for predicting protein localization sites in eukaryotic cells. Genomics **14**:897-911.

18. Otero, G. C., M. E. Harris, J. E. Donello and T. J. Hope. 1998. Leptomycin B inhibits equine infectious anemia virus Rev and feline immunodeficiency virus rev function but not the function of the hepatitis B virus posttranscriptional regulatory element. J. Virol. **72**: 7593 7597.

19. Rowland, R. R., R. Kervin, C. Kuckleburg, A. Sperlich and D. A. Benfield. 1999. The localization of porcine reproductive and respiratory syndrome virus nucleocapsid protein to the nucleolus of infected cells and identification of a potential nucleolar localization signal sequence. Virus Res. **64**:1-12.

20. Rowland, R. R., P. Schneider, Y. Fang, W. Wootton, D. Yoo, and D. A. Benfield. 2003. Peptide domains involved in the localization of the porcine reproductive and respiratory syndrome virus nucleocapsid protein to the nucleolus. Virology **16**:135-145.

21. Rowland, R. R., and D. Yoo. 2003. Nucleolar-cytoplasmic shuttling of PRRSV nucleocapsid protein: a simple case of molecular mimicry or the complex regulation by nuclear import, nucleolar localization and nuclear export signal sequences, Vir. Res.**95**: 23265 33.

22. Siomi, H., H. Shida, S. H. Nam, T. Nosaka, M. Maki, and M. Hantanaka. 1988. Sequence requirements for nucleolar localization of human T-cell leukemia virus type I pX proteins, which regulates viral RNA processing. Cell **55**:197-209.

23. Tijms, M. A., Y. van der Meer, and E. J. Snijder. 2002. Nuclear localization of non structural protein 1 and nucleocapsid protein of equine arteritis virus. J. Gen. Virol. **83**: 795271 800.

24. Wurm, T., H. Chen, T. Hodgson, P. Britton, G. Brooks, and J. A. Hiscox. 2001. Localization to the nucleolus is a common feature of coronavirus nucleoproteins, and the

protein may disrupt host cell division. J. Virol. **75**:9345-9356.

25. Yoo, D., S. K. Wootton, G. Li, C. Song and R. R. Rowland. 2003. Colocalization and interaction of the porcine arterivirus nucleocapsid protein with the small nucleolar RNA277 associated protein fibrillarin. J. Virol. **77**:12173-12183.

Section 3

An aspartic acid residue at position 372 is involved in the cytoplasmic retention of the SARS-coronavirus nucleocapsid protein

Abstract

Previous work characterizing the intracellular localization of the severe acute respiratory syndrome coronavirus (SARS-CoV) nucleocapsid (N) protein attributed the absence of nuclear localization to the inability of a 369-389 lysine-rich domain to possess a functional nuclear localization signal (NLS) sequence (Rowland et al. 2005, J. Virol. 79: 11507-11512). The capacity of mutated SARS-CoV N 369-389 peptides to restore nuclear localization to a NLS-minus equine infectious anemia virus Rev (ERev) protein was used to probe the mechanism for the absence of 369-389 NLS activity. Replacement of acidic residues with alanines at positions 372, 377 and 379 showed that a single alanine for aspartic acid substitution at position 372 restored nuclear translocation activity to the ERev-SARS N (369-389) chimeric protein. The replacement of Asp-372 with alanine in the full-length N protein resulted in the increased accumulation of GFP-tagged SARS-CoV N protein in the nucleus. A single aspartic acid as a mechanistic basis for the absence of NLS function is consistent with structural models of NLS binding to importin-α. A search for other SARS CoV N protein polypeptides with nuclear localization activity identified a potential cryptic NLS in the N-terminal region.

Introduction

Severe acute respiratory syndrome (SARS) is caused by an enveloped, positive-stranded RNA virus, the SARS coronavirus (SARS-CoV). Along with the arteriviruses, toroviruses and roniviruses, the coronaviruses are placed in a single order, *Nidovirales* (4). Even though nidovirus replication is restricted to the cytoplasm, the nucleocapsid protein (N) of two arteriviruses, porcine reproductive and respiratory syndrome virus (PRRSV) and equine arteritis virus (EAV), and as representative members of group I, II and III coronaviruses localize to the nucleolus during infection. Nuclear accumulation occurs in cells transfected with N protein genes fused to the enhanced green fluorescent protein (EGFP), demonstrating that the translocation of N across the nuclear pore complex (NPC) and accumulation in the nucleolus are independent of other viral proteins (12, 19, 20, 22, 23, 25, 27,28, 29, 32, 33). Once inside the nucleolus, the N proteins of PRRSV and infectious bronchitis virus (IBV) co-localize with major nucleolar proteins (6, 29). Presumably, it is the modulation of nucleolar protein function that represents a viral strategy that diverts biosynthetic resources from the dividing nucleus to the cytoplasm, the site of nidovirus replication (12, 13, 25, 29).

The classical mechanism for the shuttling of proteins between the cytoplasm and the nucleus is mediated by a superfamily of receptor transport proteins known as karyopherins. The karyopherins are subdivided into importins, responsible for the transport of cargo proteins into the nucleus, and exportins, which shuttle proteins from the nucleus back to the cytoplasm. Linear oligopeptide localization signal sequences are

responsible for determining the direction of transport. For instance, translocation of a cargo protein from the cytoplasm to the nucleus is initiated through the interaction between a nuclear localization signal (NLS) on the cargo protein with the NLS-binding site on importin-α (reviewed in 2 and 9). In a similar manner, the transport from the nucleus to the cytoplasm involves a nuclear export signal (NES) sequence. Classical NLS sequences are composed of short peptide regions enriched in basic amino acid residues. Monpartitie NLS sequences contain a single cluster of at least four basic residues and are further divided in pat4 and pat7 motifs. Bipartite NLS sequences contain two basic amino acid clusters separated by a 10 amino acid spacer. The NLS binding site of importin-α is formed by nine armadillo (Arm) repeats. The formation of a stable complex between the peptide NLS and the α-helices of the Arm repeats occurs through a combination of hydrophobic and electrostatic interactions.

Based on the predicted peptide sequence of the SARS-CoV N protein gene, Marra et al. (20) was the first to identify a unique lysine-rich domain located between amino acids 369-389 and suggested that the embedded nuclear/nucleolar localization signal sequences could play an important role in SARS-CoV cytopathogenesis. The identification of topogenic peptide sequence motifs using computer-based programs, such as PSORT (Prediction of protein sorting signals and localization sites in amino acid sequences; http://psort.hgc.jp/) identifies at least eight NLS motifs within the SARS-CoV N protein polypeptide, including both monopartite and bipartite NLSs. At least five NLS motifs are embedded within the 369-389 region. Even though bioinformatic analysis predicts its nuclear translocation, experimental studies of SARS-CoV N protein localization have

met with mixed results. Chang et al. (5) provided the first published report of SARS-CoV N protein in the nucleus, which was followed by Quinfen et al.(24), who observed nuclear accumulation of virus-like particles in infected Vero cells. In contrast, Timani et al. (31), detected only small quantities of N in the nucleus of infected cells and nuclear accumulation was absent when recombinant N was expressed in Vero cells as a GFP fusion protein. However, nuclear/nucleolar accumulation of recombinant N-GFP was recovered following the removal of amino acids 220 through 231, which covers a leucine-rich NES-like motif. More recently, Jaehwan et al., (2005) did not observe the presence of N in the nucleus of infected cells. However, 10% of cells transfected with a plasmid containing N fused to the enhanced cyan fluorescent protein (ECFP) gene did show nucleolar accumulation of the tagged protein. Nucleolar accumulation of N-ECFP only became apparent at 72 h after transfection. The removal of the NES-like domain failed to increase the percentage of cells with nuclear/nucleolar localization. Surjit et al., studying N protein phosphorylation, reported the presence of phosphorylated N in the nucleus, suggesting that the intracellular localization of N is determined by the phosphorylation state of the N protein (29). We failed to observe N in the nucleus following the infection of Vero cells with the SARS-CoV Urbani isolate. Similar results were obtained after the transfection of Vero and other cell lines with plasmid constructs that expressed N as an EGFP fusion protein. Accumulation of the EGFP tag in the nucleus of some cells was attributed to the proteolytic removal of EGFP. The absence of N protein nuclear localization was linked to the inability of the 369-389 region to possess a functional NLS. The lack of NLS activity within the 369-389 oligopeptide is intriguing, since the 369-PKKDKKK peptide sequence is similar to other well-studied pat7 NLS

sequences, including the NLS (126-PKKKRK) of the SV40 large T antigen and the 41-PGKKNKK peptide sequence in the PRRSV N protein. In addition, the SARS CoV NLS 369 pat7 motif is nearly identical to the putative NLS, 360-PKKEKKL, reported for the IBV N protein (13). It should be noted that the identification of the IBV N 360-366 peptide as a functional NLS has yet to be confirmed experimentally.

NLS function is partially dependent on the capacity of key lysine/arginine residues to form electrostatic bonds with corresponding negatively charged residues located within the NLS binding pocket of importin-a. We hypothesized that negatively charged residues located within the 369-389 peptide could interfere with NLS binding. Substituting lysine residues at 372, 378 and 379 identified an aspartic acid residue at position 372, which when replaced with an uncharged alanine residue was sufficient to restore NLS activity to the 369-389 peptide. Therefore, the presence of an aspartic acid at position 372 plays an important role in the retention of the N protein in the cytoplasm. A role for the lysine-rich domain in the pathogenesis of SARS-CoV remains to be determined.

MATERIALS AND METHODS

Cells and reagents Vero cells were grown in Eagle minimum essential medium supplemented with 10% fetal bovine serum, penicillin-streptomycin and fungizone. Cells were maintained at 37^{o}C in 5% CO_2. Leptomycin B (LMB) was obtained from Sigma and stored at -20 until use.

Construction of recombinant plasmids The cDNA representing the N gene of the Urbani isolate of SARS-CoV was a generous gift of Paul Rota at CDC. PCR was used to place EcoRI and BamHI restriction sites on the 5' and 3' ends of the N gene cDNA, respectively. The cDNA was cut with the restriction enzymes and cloned into pEGFP-C2. The product was a plasmid that expressed the N protein fused to the C-terminus of EGFP. Expression of the N-EGFP fusion protein was regulated by the CMV immediate early promoter. Truncated N gene cDNAs expressing different regions of the N protein were constructed using the same approach. All N gene constructs were confirmed by DNA sequencing. The plasmid construct, pEGFP-ERev, which contained the full-length cDNA of the EIAV Rev protein (ERev) fused to EGFP, was a kind gift of Susan Carpenter. The NLS of ERev was removed by deleting 24 bases from the 3' end of the full-length ERev cDNA, producing the construct, pEGFP-ERev(NLS-). For transfection, Vero cells were plated onto 24-well tissue culture plates. Cells, when 75% confluent, were transfected with 2ug of plasmid DNA using lipofectamine 2000 (Invitrogen). Four hours after transfection, medium was replaced by maintenance medium and visualized under fluorescence microscope (Nikon) 15-18 h later. For LMB treated cells, LMB was added at concentrations of 10 nM at between 3-4 h after transfection.

Immunostaining and confocal microscopy Confocal microscopy was performed on transfected Vero cells grown on coverslips. Cells were fixed for 10 minutes with 2% paraformaldehyde in PBS, washed and counterstained with TO-PRO-3 (Molecular Probes) and then viewed on a Zeiss LSM 510 confocal microscope (). For the detection of N using an anti-N mAb, cells were fixed in 80% acetone for 10 minutes and then stained with SA 46-4 anti SARS N mAb. The mouse mAb was detected goat anti-mouse IgG antibody conjugated to AlexaFluor 594 (Molecular Probes). All antibodies were diluted 1:500 in PBS with 5% FBS and incubations performed for 1 h at room temperature followed by extensive washing in PBS. Cells were counterstained with TO-PRO-3, diluted in blocking buffer, then viewed on a Zeiss LSM 510 confocal microscope.

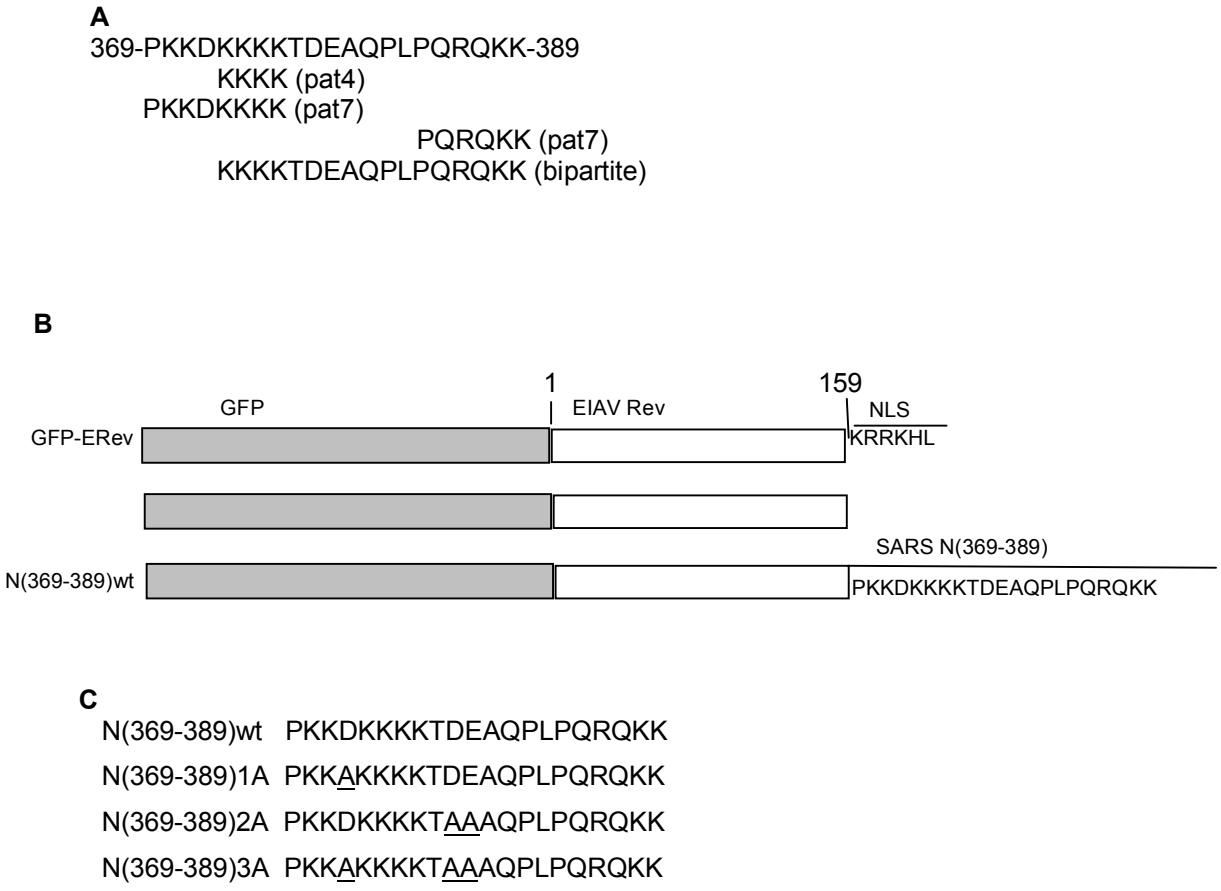


Fig. 3.1. NLS motifs in SARS-CoV N protein 369-389 peptide.

(A) Peptide sequence analysis was performed using the Web-based program, PSORT (Nakai and Kanehisa, 1992). The "pat4" NLS is defined as a continuous stretch of four basic amino acids (lysine or arginine) or three basic amino acids associated with a histidine or proline. The "pat7" NLS motif starts with a proline and is followed within three residues by a segment containing three basic residues out of four. The "bipartite" motif, consists of two basic amino acids, a 10 amino acid spacer, and a five amino acid segment containing at least three basic residues. (B) Wild-type and mutated EGFP-ERev

constructs used in this study. (C) Mutant 369-389 peptides containing alanine substitutions (identified by underlined residues).

RESULTS

Effect of negatively charged amino acid residues on the nuclear localization activity of the SARS-N(369-389) peptide. As shown in Fig. 3.1A, the 369-389 region of SARS N possesses multiple NLS motifs, including pat4, pat7 and bipartite signal sequences. In a previous study we tested the nuclear transport activity of the SARS-N (369-389) peptide using a cis complementation approach to recover nuclear localization of a NLS-minus equine infectious anemia virus Rev (ERev) protein. ERev possesses a single monopartite NLS, KRRRK, located at the C-terminal end, which is necessary and sufficient to import the 165 amino acid ERev protein into the nucleus. A single, atypical NES, 32-PQGPLESDQWCRVLRQSLPEEKIP, is responsible for nuclear export (18, 21). Under steady-state conditions, the net activities of the nuclear import and export signal sequences favor the accumulation of recombinant EGFP-ERev in the cytoplasm (Fig. 3.2A). After the addition of the CRM1 exportin inhibitor, LMB, nuclear export is blocked and the distribution shifts towards the accumulation of EGFP-ERev in the nucleus (Fig.3.2B). To study the nuclear import function of the SARS-N (369-389) peptide, we removed the NLS of EGFP-ERev by deleting 24 bases from the 3' end of the ERev full-length cDNA (see Fig. 3.1B). The truncated construct, pEGFP-Rev (NLS-), when transfected into Vero cells expressed a GFP-labeled ERev protein that was retained in the cytoplasm, even in the presence of LMB (Fig. 3.2C and D). The fusion of the SARS-N (369-389) peptide to the C-terminus of GFP-ERev(NLS-); see construct N(369-389)wt in Fig 3.1B), did not alter the localization properties of the GFP-ERev(NLS-) protein (see Fig. 3.2E and F). This result, as reported by us previously, indicates that the 369-389 peptide lacks a functional NLS (28)

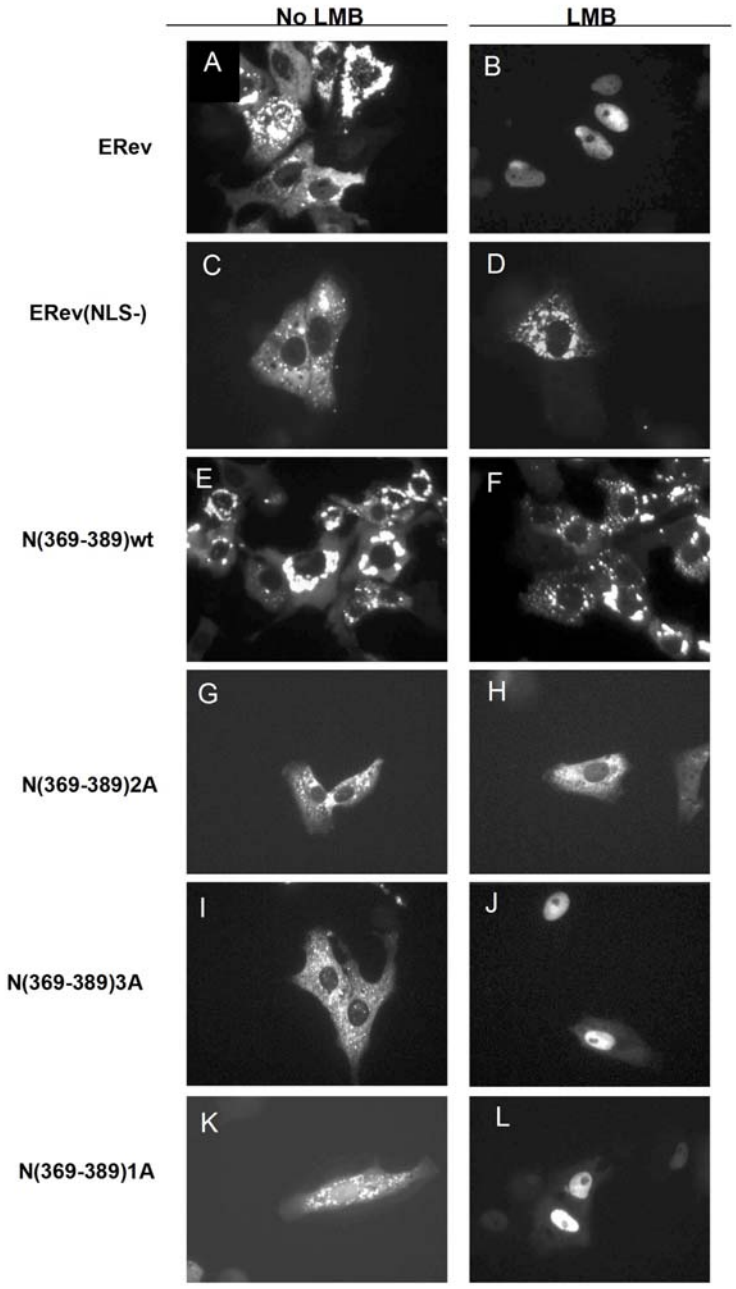

No LMB
LMB
A
B
ERev
C
D
ERev(NLS-)
E
F
N(369-389)wt
G
H
N(369-389)2A
I
J
N(369-389)3A
K
L
N(369-389)1A

Fig.3.2. Localization properties of EGFP-ERev-SARS-N chimeric proteins. Photomicrographs of live cells transfected with pEGFP-ERev constructs. Right hand panels are cells treated with LMB. LMB, at a concentration of 10 nM, was added at 4h after transfection. Photomicrographs were taken at 18 h after transfection.

One explanation for the lack of NLS activity is the presence of several negatively charged amino acid residues, located in and around the NLS motif 369-PKKDKKK. The effect of these negative charges would be to either neutralize the positive charges conferred by the nearby lysine residues, or to alter the conformation or accessibility of the NLS. Acidic residues are found at positions 372, 378 and 379. An additional source for negative charges is found in the post-translational phosphorylation of serine, threonine and tyrosine residues by cellular kinases. A single threonine, located at position 377, was identified by the computer program, NetPhos (http://www.cbs.dtu.dk/services/NetPhos/; 3), as a potential target for phosphorylation. We replaced the aspartic and glutamic acids within the 369-389 peptide with uncharged alanine residues and tested the ability of the mutated peptides to restore nuclear localization to GFP-ERev (NLS-). The peptide sequences of wild-type and mutated 369-389 constructs are shown in Fig. 3.1C. The replacement of Asp-378 and Glu-379 with alanines, in the construct SARS-N(369-389)2A, did not alter the localization properties of the EGFP-ERev(NLS-). Even in the presence of LMB, the chimeric protein was retained in the cytoplasm (see Fig. 3.2G and H). The further substitution of Asp-372 with alanine produced the triple mutant, SARS-N(369-389)3A. In the absence of LMB, the GFP-labeled protein was predominately

cytoplasmic (Fig. 2I). However, in the presence of LMB, the GFP-tagged protein localized almost entirely to the nucleus (Fig. 3.2J). Therefore, we prepared a third construct, pSARS-N(369-389)1A, which contained a single alanine substitution at position 372 (Fig. 3.1C). The localization properties of the SARS-N(369-389)1A protein, with and without LMB, were similar to the triple-mutant (Fig. 3.2K and L). These results indicate that the replacement of Asp-372 with an alanine is sufficient to restore NLS activity to the 369-389 peptide. The replacement of Thr-377 with alanine did not alter the localization properties of the ERev(NLS-) protein (data not shown).

Effect of alanine-372 substitution on SARS-CoV N protein localization: In order to determine if Asp-372 was important for retaining SARS N in the cytoplasm, we used site-directed mutagenesis to replace Asp-372 with alanine within the full length SARS N protein. This was accomplished by replacing the aspartic acid codon sequence, GAC, with the alanine codon, GCC. The resulting construct was named pEGFP-SARS-N (372A). When screening plasmid constructs for the presence of the Ala-372 mutation, we identified a double mutant, which possessed an additional methionine for lysine substitution at the adjacent 371 position. This mutation was the result of a change in the lysine codon, AAG, to methionine, AUG. This resulting construct was named pEGFP-SARS-N (371M,372A). Duplicate wells of Vero cells were transfected with pEGFP-SARS-N, pEGFP-SARS-N (372A) or pEGFP-SARS-N(371M,372A). At three hr after transfection, one well of each construct was treated with LMB and 15 h later the live cells were viewed under a fluorescence microscope. The rationale for the inclusion of LMB was based on a report by Timani et al. (31), who identified a leucine-rich NES-like

peptide sequence, 220-LALLLLDRLNRL, located in the middle of the SARS-CoV N protein.

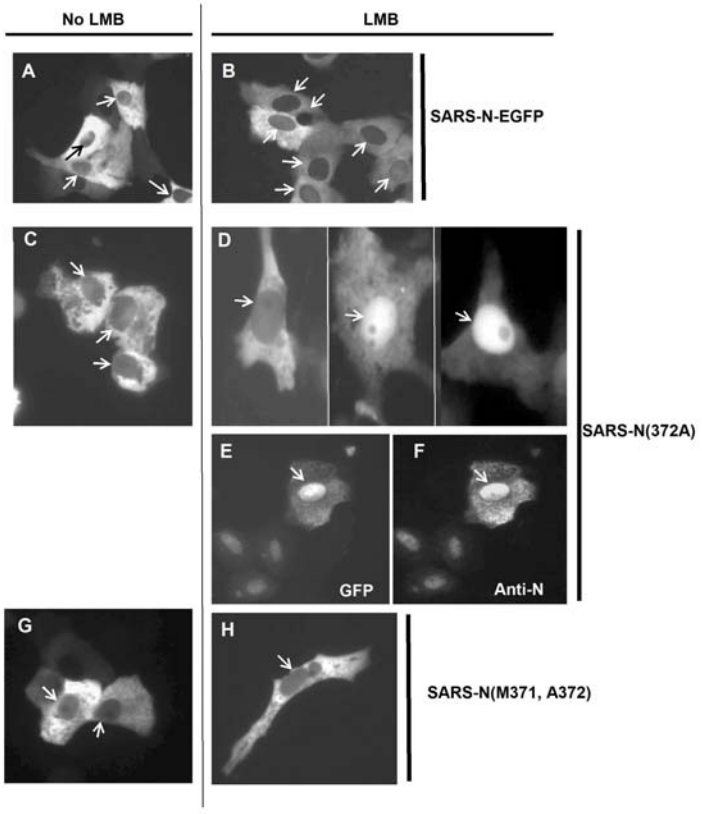


Fig. 3.3. Localization properties of mutated EGFP-SARS-N proteins. Right hand panels are the results following treatment of transfected cells with LMB. Images are of live cells except E and F, which are the same cell after fixation prior to imaging. (E) shows EGFP green fluorescence and (F) is after staining with anti SARS-Co-N mAb, 46-4 followed by AlexaFluor 594 and viewed under a rhodamine filter. Arrows identify the location of the nucleus.

As reported by us previously, the full-length SARS-N protein when fused to the C-terminal end of EGFP is retained in the cytoplasm, even in the presence of LMB (see Fig. 3.3A and B). The localization properties of the EGFP-SARS-N (372A) protein are shown in Fig. 3.3C and D. The EGFP-SARS-N (372A) protein was predominately cytoplasmic, with a small number of cells with EGFP in the nucleus. In the presence of LMB, we observed a heterogeneous population of fluorescent cells, ranging from cells that retained EGFP in the cytoplasm to cells with EGFP exclusively in the nucleus. Examples of the various localization patterns in LMB-treated cells are shown in Fig. 3D. One explanation for the accumulation of EGFP in the nucleus is the result of proteolytic removal of SARS-N (372A) EGFP and the passive diffusion of free EGFP across the NPC and into the nucleoplasm. Staining of cells with the anti-N antibody, 46-4, showed that 46-4 staining co-localized with EGFP in the nucleus of LMB-treated cells (see Fig. 3.3F and G). These results indicated that similar to the results for the ERev-SARS-N (368-389) chimeric protein, the presence of an alanine at position 372 restored nuclear translocation to the SARS-CoV N protein. These results also suggest that SARS-N possesses a NES-like domain. The localization properties of the SARS-N (371M, 372A) double mutant resembled the wild-type N protein (see Fig. 3.3G and H). The absence of nuclear accumulation is consistent with the requirement of Lys-371 for NLS function.

A more quantitative assessment of the effect of LMB on the localization properties of wild-type and mutant N proteins is presented in Fig. 3.4. In this experiment we determined the percentage of cells with fluorescence in nuclear and cytoplasmic compartments. Data for EGFP-ERev, shown in Fig. 3.4A were included as a positive

control. After transfection with pEGFP-ERev, approximately 60% cells contained EGFP-ERev only in the cytoplasmic compartment. In the presence of LMB the distribution was shifted to almost 100% of cells with EGFP-tagged protein in the nucleus. Fig. 3.4B shows that approximately 80% of cells expressing EGFP-labeled SARS-N exhibited a predominately cytoplasmic distribution, which did not change in cells incubated in the presence of LMB. Results for cells transfected with pSARS-N (372A) are shown in Fig. 3.4 C. In the absence of LMB, EGFP-SARS-N (372A) cells with a predominately cytoplasmic distribution were dominant, with only a small shift towards an increased percentage of cells with fluorescence in the nucleus (compare open bars in panels B and C in Fig. 3.4). In response to LMB, the distribution was further shifted to the right, reflecting an increased percentage of cells with fluorescence in the nucleus. However, the shift was not as pronounced as the LMB response of EGFP-ERev cells. The results for the EGFP-SARS-N (371M, 372A) double mutant were similar to wild-type N (see panel D). Collectively, these results indicate that a single aspartic acid residue at position 372 contributes to the cytoplasmic retention of SARS-CoV N protein.

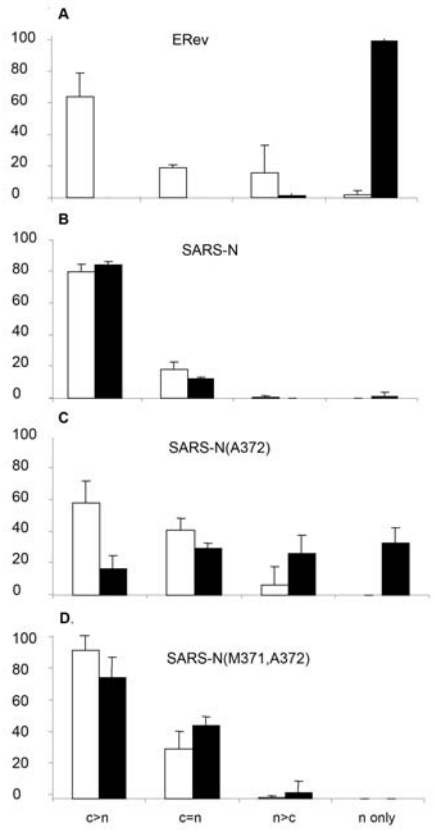


Fig. 3.4. Intracellular distribution of EGFP fluorescence in live cells. Data for each experiment was collected by visually assessing the intracellular distribution of fluorescence in Vero cells in at least five microscopic fields. Results show means and standard deviations for data from three independent experiments. Open bars are control cells and closed bars after treatment with 10 nM LMB. Vertical axis shows percent cells. Key: C>N, fluorescence greater in cytoplasm than nucleus; C=N, fluorescence in cytoplasm equal to fluorescence in nucleus; C<N, fluorescence greater in greater in nucleus than cytoplasm; N=only, fluorescence on in the nucleus.

The N-terminal region in SARS N possesses nuclear localization activity. In addition to the NLS motifs embedded within the 369-389 region, PSORT identified two other regions in SARS-N possessing NLS motifs. The first, called NLS-1, contained the pat7 motif, PKQRRPQ, located between amino acids 38-44. The second NLS motif, called NLS-2 possessed overlapping NLS motifs; a pat4 peptide, KKPR, between amino acids 257 and 260 and a pat7 motif, 259-PRQKRTA. And the third, labeled NLS-3, contained the 369-389 peptide. We prepared several N protein gene cDNA constructs containing different regions of the N protein fused to EGFP. The constructs and their nuclear localizaton properties are shown in Fig. 3.5A. Only two constructs, pEGFP-SARS-N(1-100) and pEGFP-SARS-N(1-200), showed accumulation of the EGFP tag in the nucleus. Confocal microscopy of a representative cell transfected with pEGFP-SARS-N(1-100), is presented in Fig. 3.5B. The green EGFP-tagged protein accumulated in the same compartment as the blue TO-PRP-3 nuclear stain. The only exception was the region covered by the nucleolus. EGFP fluorescence was present in the nucleolus; whereas, TO-PRO-3 was excluded. The same results were obtained for cells transfected with pEGFP-SARS-N-1-200. Constructs that contained NLS-1 plus NLS-2, NLS-2 alone, NLS-2 plus NLS-3, or NLS-3 alone, remained cytoplasmic.

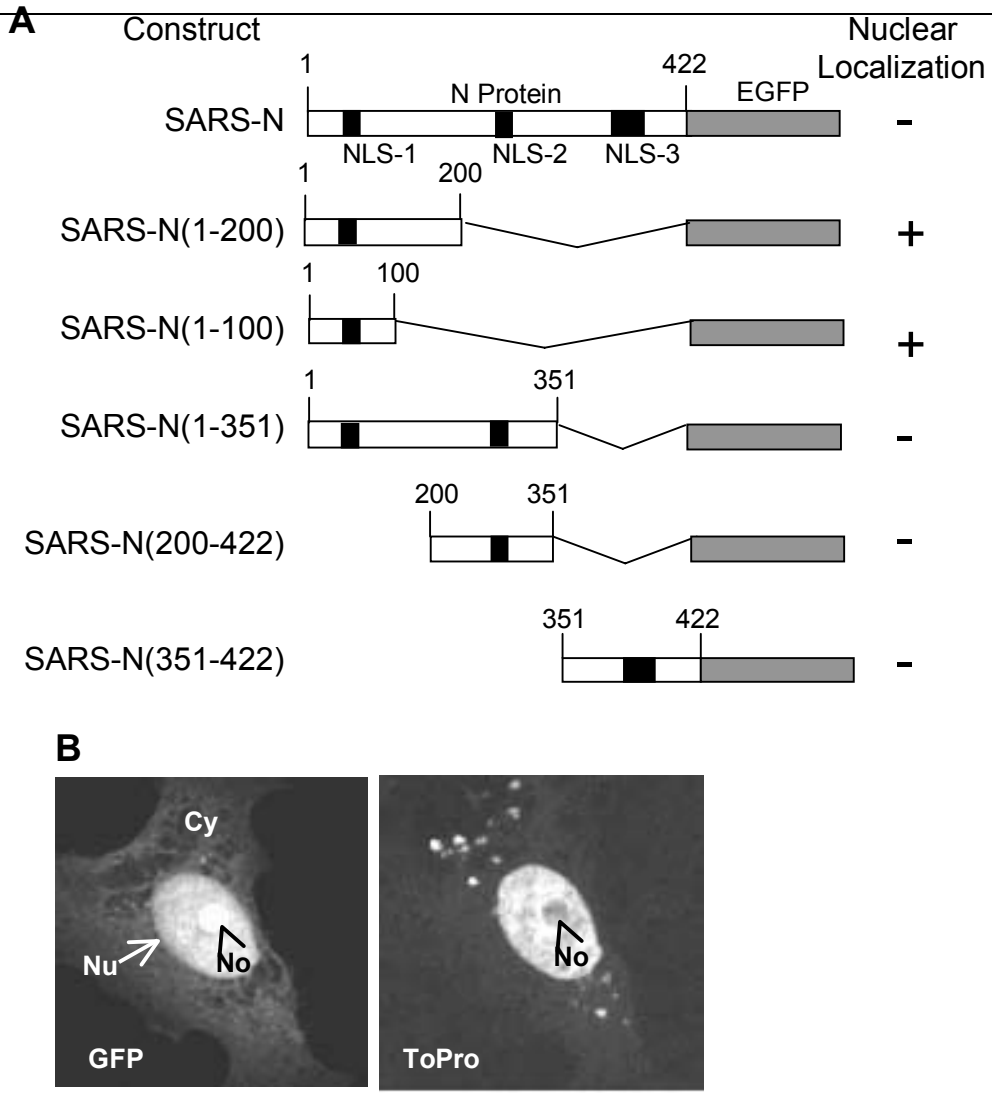


Fig. 3.5. Intracellular localization of EGFP-tagged SARS-CoV N protein fragments. (A) SARS-CoV N protein constructs are shown on the left . On the right of each construct indicates the presence (+) or absence (-) of localization to the nucleus. Results were obtained after visualizing live cells at 18-24 hr following transfection of each plasmid in Vero cells. (B) Confocal microscopy of a representative cell transfected with pEGFP-SARS-N(1-100). Left image is EGFP fluorescence. The right image shows staining with the nuclear stain, TO-PRO-3 , in the same cell. Key: Cy, cytoplasm; Nu, nucleus; No, nucleolus.

Discussion

The nuclear localization of the nucleocapsid to the nucleus/nucleolus is a common property within the nidovirus group. At present, the only exception appears to be the N protein of SARS-CoV, which possesses at least eight canonical NLS motifs, and yet is retained in the cytoplasm. In previous work, we attributed the cytoplasmic retention of the SARS CoV to the inability of a NLS motif -rich region, between amino acids 369-389, to possess nuclear localization activity. In this study we applied a cis complementation approach incorporating a NLS-minus ERev protein to determine the mechanistic bases for lack of nuclear localization. The replacement of a single Asp-372 with alanine was sufficient to restore localization activity to the 369-389 peptide. NLS activity was evident by the accumulation of the ERev (NLS-) SARS-N(369-389) chimeric protein in the nucleus. Further support for the importance of Asp-372 was found when Asp-372 was replaced with Alanine in the native protein. Again the accumulation of EGFP-SARS-N (372A) protein was evident after the addition of LMB. However, the effect of LMB was not as pronounced compared to ERev.

References

1. **Almazan F., C. Galan and L. Enjuanes.** 2004. The nucleoprotein is required for efficient Coronavirus genome replication. J Virol. **78**:12683-12688.
2. **Baric R., G. Nelson, J. Fleming, R. Deans, J. Keck, N. Casteel and S. Stohlman.**1988. Interactions between coronavirus nucleocapsid protein and viral RNAs: implications for viral transcription. J Virol. **62**:4280-4287.
3. **Blom N**, **Gammeltoft S**, **Brunak S**. 1999. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. J Mol Biol. Dec 17;**294(5):**1351-62.
4. **Cavanagh D.**1997. Nidovirales: a new order comprising Coronaviridae and Arteriviridae, Arch. Virol. **142**:629-633
5. **Chang MS et al.** 2004. Antibody detection of SARS-CoV spike and nucleocapsid protein. Biochem Biophys Res Commun. 2004 Feb **20**;314(4):931-6.
6. **Chen H., T. Wurm, P. Britton, G. Brooks and J.A. Hiscox.** 2002. Interaction of the coronavirus nucleoprotein with nucleolar antigens and the host cell, J Virol. **76**:5233-5250.
7. **Fontes M., T. Teh and B. Kobe.** 2000. Structural basis of recognition of monopartite and bipartite nuclear localization sequence by mammalian importin-α. J Mol. Biol**. 297**:1183-1194.
8. **Fornerod M., M. Ohno, M. Yoshida, and I. Mattaj.** 1997. CRM1 is an export receptor for leucine-rich nuclear export signals. Cell. **90**:1051-1060.

9. **Gorlich D. and U. Kutay.** 1999. Transport between the cell nucleus and the cytoplasm. Ann. Rev. Cell. Dev. Biol. **5**:607-660.

10. **He R., A. Leeson , M. Ballantine , A. Andonov , L. Baker , F. Dobie , Y. Li , N. Bastien, H. Feldmann, U. Strocher , S. Theriault, T. Cutts, J. Cao , T.F. Booth, F.A Plummer, S. Tyler and X. Li.** 2004. Characterization of protein-protein interactions between the nucleocapsid protein and membrane protein of the SARS coronavirus. Virus Res. **105**:121-125.

11. **Hicks G.R. and N.V. Raikhel.** 1995. Protein import into the nucleus: an integrated view. Ann. Rev. Cell Dev. Biol. **11**:155-188.

12. **Hiscox J. A.** 2002. The nucleolus--a gateway to viral infection? Arch. Virol.**147**:1077- 1089.

13. **Hiscox J.A., T. Wurm, L. Wilson, P. Britton, D. Cavanagh and G. Brooks.** 2001. The coronavirus infectious bronchitis virus nucleoprotein localizes to the nucleolus, J Virol. **75**:506-512.

14. **Huang Q., L. Yu , A.M. Petros, A. Gunasekera , Z. Liu , N. Xu , P. Hajduk, J.Mack , S.W. Fesik and E.T. Olejniczak .** 2004. Structure of the N-terminal RNA-binding domain of the SARS CoV nucleocapsid protein. Biochemistry. **43:**6059-6063.

15. **Jans D.A., and S. Hubner.** 1996. Regulation of protein transport to the nucleus: central role of phosphorylation. Physiol. Rev. **76**:651-685.

16. **Jans D.A., C.Y. Xiao and M.H Lam**. 2000. Nuclear targeting signal recognition: a key control point in nuclear transport? Bioessays. **22**:532-544.

17. **Knipe D. and P. Howley.** 2001. Fields Virology. Fourth Edition. Lippincott Williams and Wilkins.

18. **Kubota S., H. Siomi, T. Satoh, S. Endo, M. Maki, and M. Hatanaka.** 1989. Functional similarity of HIV-I rev and HTLV-I rex proteins: Identification of a new nucleolar targeting signal in rev protein. Biochem. Biophys. Res. Comm. **162**:963-970.

19. **Kubota S., T. Copeland and R. Pomerantz.** 1999. Nuclear and nucleolar targeting of human ribosomal protein S25: Common features shared with HIV-1 regulatory proteins. Oncogenes. **18**:1503-1514

20. **Marra MA et al.** 2003. The Genome sequence of the SARS-associated coronavirus. Science. 2003 May **30**;300(5624):1399-404. Epub 2003 May 1

21. **Meyer B. E., J. L. Meinkoth and M. H. Malim.** 1996. Nuclear transport of human immunodeficiency virus type 1, visna virus, and equine infectious anemia virus Rev proteins: identification of a family of transferable nuclear export signals. J Virol. **70**:2350-2359.

22. **Nakai K. and M. Kanehisa.** 1992. A knowledge base for predicting protein localization sites in eukaryotic cells. Genomics. **14**:897-911.

23. **Otero, G. C., M. E. Harris, J. E. Donello and T. J. Hope.** 1998. Leptomycin B inhibits equine infectious anemia virus Rev and feline immunodeficiency virus rev function but not the function of the hepatitis B virus posttranscriptional regulatory element. J Virol.**72**:7593- 7597.

24. **Qinfen Z, Jinming C, Xiaojun H, Huanying Z, Jicheng H, Ling F, Kunpeng L, Jingqiang Z.** 2004. The life cycle of SARS coronavirus in Vero E6 cells. J Med Virol.Jul;**73(3):**332-7.

25. **Rowland R.R., D. Yoo.** 2003 b. Nucleolar-cytoplasmic shuttling of PRRSV nucleocapsid protein: a simple case of molecular mimicry or the complex regulation by nuclear import, nucleolar localization and nuclear export signal sequences. Virus Res. **95:**23-33

26. **Rowland R.R., P. Schneider, Y. Fang, S. Wootton, D. Yoo and D.A. Benfield.** 2003 a. Peptide domains involved in the localization of the porcine reproductive and respiratory syndrome virus nucleocapsid protein to the nucleolus. Virology. **316:**135-145.

27. **Rowland R.R., R. Kervin, C. Kuckleburg, A. Sperlich, D.A Benfield.** 1999. The localization of porcine reproductive and respiratory syndrome virus nucleocapsid protein to the nucleolus of infected cells and identification of a potential nucleolar localization signal sequence. Virus Res. **64:**1-12.

28. **Rowland R.R., V. Chauhan, Y. Fang, A. Pekosz, M. Kerrigan and M. Burton.** 2005. Intracellular localization of the SARS coronavirus nucleocapsid protein: absence of nucleolar accumulation during infection and after expression as a recombinant protein in Vero cells. J Virol.**79**:11507-11512.

29. **Surjit M., B. Liu, P. Kumar, V.T Chow and S.K. Lal.** 2004. The nucleocapsid protein of the SARS coronavirus is capable of self-association through a C-terminal 209 amino acid interaction domain. Biochem Biophys Res Commun. **317**:1030-1036

30. **Tijms M.A., Y. Van Der Meer and E.J. Snijder.** 2002. Nuclear localization of non-structural protein 1 and nucleocapsid protein of equine arteritis virus. J Gen Virol. **83**:795-800.

31. **Timani KA**, **Liao Q**, **Ye L**, **Zeng Y**, **Liu J**, **Zheng Y**, **Ye L**, **Yang X**, **Lingbao K**, **Gao J**, **Zhu Y**. 2005. Nuclear/nucleolar localization properties of C-terminal nucleocapsid protein of SARS coronavirus. Virus Res.Dec;**114 (1-2):**23-34. Epub 2005 Jun 29.

32. **Wurm T., H. Chen, T. Hodgson, P. Britton, G. Brooks and J.A Hiscox.** 2001. Localization to the nucleolus is a common feature of coronavirus nucleoproteins, and the protein may disrupt host cell division. J Virol. **75**:9345-9356.

33. **Yoo D., S.K. Wootton, G. Li, C. Song and R.R. Rowland.** 2003. Colocalization and interaction of the porcine arterivirus nucleocapsid protein with the small nucleolar RNA-associated protein fibrillarin, J Virol. **77:**12173-12183.

Section IV

Conclusion and Future Directions

SARS coronavirus is neither believed to be a mutant form not a recombinant of any known human coronaviruses. Serological tests have determined Chinese Horse-shoe bats to be the original reservoirs and virus has obtained the ability to infect humans. Like many othercoronaviruses, SARS-CoV is spread by both fecal and respiratory contamination. Development of vaccines and anti-viral therapy for SARS-CoV will provide new strategies for other animal and human coronaviruses.

The nucleocapsid protein of the coronaviruses is a highly immunogenic protein and this property is true for SARS-CoV as well. More than 90% of infected patients show antibodies against this virus. The N- protein plays a vital role in viral core formation, pathogenesis and transcription. These properties make the nucleocapsid an excellent target for vaccine development. Antibodies generated against this protein are used for diagnostic purposes.

The research documented in this dissertation is an effort to study the role of the N in viral pathogenesis of SARS-CoV. Studying the localization pattern of the viral protein upon infection is the first step to study its cellular interactions and association with other molecules. Although the protein possessed multiple localization signals as predicted by PSORT, localization was restricted to the cytoplasm of the (Section 2). Even though the various predicted NLSs were similar to the consensus NLS domains, the sequences do not serve as active localization domains. The fragment encoding amino acids 369-389 showed nuclear localization potential. True localization signals possess two very important properties. Addition of these signals can localize a non-nuclear protein to the

nucleus and a single amino acid mutation in this region results in the loss of nuclear targeting property. To test the true localization potential of the 369-389 fragment, we attached this fragment to a NLS lacking EIAV-Rev protein. This recombinant protein thus was not targeted to the nucleus of the transfected cell (section 2). Thus, the 369-389 fragment even though capable of localizing to the nucleolus by passive diffusion does not participate in active transport and does not target the protein to the nucleus of the cell.

To further understand the potential bipartite signal domain, we analyzed the amino acids in this fragment. The results of this study are shown in section 3. This work reflects on the role of the aspartic acid at the position 372 on the cytoplasmic retention of the protein. Nuclear localization signals are basic amino acids and they interact with the acidic aspartic and glutamic amino acids found in the ARM motifs of importin-α. We believe that the acidic aspartic residue makes the lysine rich domain incompatible to interact with the importin-α interaction regions. To study this possibility, single amino acid mutations of the three acidic amino acids were generated. The mutations at position 378 and 379 did not alter the localization properties. However, a single point mutation at position 372 changed the localization of the recombinant protein. Upon addition of LMB, the protein retention in the nuclear matrix increased. Additionally, we established the nuclear localization ability of SARS-N proteins N-terminal.

The novel localization pattern of the nucleocapsid protein conceivably reflects on unique characteristics and mechanisms of this protein. The N protein is the most abundant viral protein found in infected cells. Further studies will reflect on the functions and

mechanism of action of this molecule and define its role in viral propagation and pathogenesis. Fuller understanding of the protein conformation and molecular interactions will help in establishing the functional aspect of this protein, and this in turn will be instrumental in studying therapeutic targets for virus treatment.