4 Discussion

4.1 Emergence of SARS-CoV and susceptibility studies

4.1.1 Coronavirus antibodies in African bat species

Recently, SARS-like coronaviruses (SL-CoVs) were identified in Asian bats, predominantly in the *Rhinolophus* species (Lau et al., 2005; Li et al., 2005; Ren et al., 2006; Wong et al., 2006). It was shown that coronaviruses from bats have the most genetic diversity and are older than all coronaviruses from any other animal species. It seems that these viruses are endemic in bat populations whereas in other animals there is epidemic occurrence indicating that bats are the natural reservoir for currently known coronaviruses (Vijaykrishna et al., 2007). Up to now only Asian bats have been thoroughly investigated for being the potential natural reservoir of SARS-CoV, according to the fact that the virus emerged from Guangdong province in the People´s Republic of China late in 2002 (Drosten et al., 2003; Ksiazek et al., 2003).

The presented study is the first to analyze African bat species for antibodies reactive with SARS-CoV antigen by different serological assays. Overall, 7% seropositive bats in seven out of 26 species from four different locations in two countries were identified. The percentage of positives is similar to studies in Asia where 6% of tested bats were positive for coronaviruses. As in those studies coronavirus nucleic acids could be isolated from rectal swabs, genetic diversity studies could be performed. The majority of detected coronaviruses clustered phylogenetically in group I, followed by putative group V, a novel, yet bat specific group. Only approximately 5% of detected coronaviruses belonged to putative group IV and were thus closely related to SARS-CoV (Tang et al., 2006).

The identification of coronavirus nucleic acids failed in the presented study. The negative findings obtained in RT-PCR can be explained by the unlikelihood of finding virus nucleic acid in sera. Putitavely, the virus persists in the enteric tract as observed for other animal coronaviruses and it may not be found in serum at all (Saif, 2004). Thus genetic diversity studies could not be performed as it was done in Asia. In those studies it was found out that bat coronaviruses seem to demonstrate a high degree of host restriction as most bats from the same species but from different locations carried closely related coronaviruses. Despite the

identification of coronaviruses which are closely related to SARS-CoV a direct progenitor has not been found.

Two of the main African bat species, namely *Rousettus aegyptiacus* and *Mops condylurus,* which showed high prevalence to SARS-CoV antigen are widely distributed but vary in the degree of contact with humans in Africa. *Rousettus aegyptiacus* roosts in caves but forages in orchards at night time (Cavaleros et al., 2003) whereas *Mops condylurus* roosts in buildings (personal communication Prof. Dr. Robert Swanepoel, NICD, RSA). Interestingly, and in contrast to the Asian results, the *Rhinolophus* species did not show high antibody prevalences for SARS-CoV antigen. The reason for this discrepancy could be the host restriction mentioned above, so that *Rhinolophus* specific coronaviruses might not have reached the African continent yet.

For screening purposes a commercially available ELISA which was found to be highly sensitive and specific for detection of SARS-CoV antibodies was applied. The results of the WB analyses supported the specificity of the ELISA used in this study. The IIFT is known to be less sensitive than ELISA, but still provided confirmation in a third of the tested sera. As ELISA titers of bat sera were relatively low (below 1:800) the minor sensitivity of the IIFT obviously resulted in less positive samples. One has to consider that all sera had been stored at -80°C up to 20 years with a few freeze and thaw cycles. A certain loss of quality can thus be expected. Moreover, it is likely that reactive bat antibodies have less affinity to SARS-CoV antigen epitopes as they probably result from an infection with a SARS-related bat coronavirus most likely from putative group IV. Those SARS-CoV related viruses might have different antigenic properties.

This hypothesis is emphasized by the negative virus neutralization test (VNT) results. The inability to neutralize SARS-CoV (strain Hong Kong) is not unexpected since this assay detects only antibodies which interfere with the specific entry mechanism of SARS-CoV by recognizing particular epitopes within the S protein. In several studies it could be shown that the receptor binding domain of the S protein at aa position 318 to 510 contains multiple conformation-dependent neutralizing epitopes serving as the major target of SARS-CoV neutralization (Berry et al., 2004; He et al., 2005a; He et al., 2005b; Tripp et al., 2005; He et al., 2006; ter Meulen et al., 2006). By alignment studies deletions and mutations were identified in Asian bat SL-CoV isolates within this S protein region most likely affecting the

conformation of epitopes. Therefore, cross-neutralization might not always be possible as emphasized by conflicting VNT results obtained in Asia (Lau et al., 2005; Li et al., 2005; Sui et al., 2005).

Most ELISA positive sera from the African bat collection showed reactivity in WB analysis using the S protein fragment that contains essential parts for neutralization. As in this study denaturing conditions were used, the resulting linearized epitope is no adequate substitute for evaluating neutralizing ability of conformational epitopes. In fact, others have already observed that neutralizing antibodies are not necessarily reactive in WB analysis and vice versa (Berry et al., 2004).

Certainly, a different situation is observed in Asia where population densities are high, live animal food markets are common and thus putative transmission of viruses (e.g. new influenza strains) to humans is more likely (Smith et al., 2006). Nonetheless, the results of this study suggest that some of the African bat species harbour agents related to putative group IV CoV and therefore justify further investigations to determine potential public health risks.

4.1.2 Susceptibility of different cell lines to SARS-CoV infection

In order to study and characterize newly emerged viruses, cell culture and animal models are a necessity. The development of diagnostic tools and antiviral therapies are also dependent on cell culture propagation of viruses. The identification of susceptible cell lines also gives valuable information on tissue and organ specificity of a virus and can support pathological findings. Therefore, 23 different cell lines were tested for their susceptibility to SARS-CoV.

In this present study it could be demonstrated that the green monkey cell line Vero E6 and the human hepatoma Huh-7 cells show a high susceptibility to SARS-CoV resulting in high infection rates within hours. By using the quantitative real-time PCR up to 6 x 10^7 viral RNA copies in 2.5 x 10^4 SARS-CoV infected Vero E6 cells could be detected 31 h after infection. These data could be confirmed with IFA showing up to 100% infection of Vero E6 cells and 50% infection of Huh-7 cells 31 h after infection. Another finding was the demonstration that SARS-CoV could replicate in porcine PS and POEK cells. These observations have also been made by others who detected replication of SARS-CoV in the porcine cell line PK-15

(Giroglou et al., 2004). However, in the presented experiments infection rates of the porcine cells with SARS-CoV were clearly lower and resulted in a putative persisting infection. Possible reasons for lower infection rates are discussed below. Pigs, as domestic animals on Chinese markets, have been discussed as putative amplifying hosts for SARS-CoV but could apparently not be infected (Weingartl et al., 2004).

The angiotensin converting enzyme 2 (ACE2) has been identified to play an important role in SARS-CoV entry (Li et al., 2003). It can be expected that the SARS-CoV susceptible cells express a SARS-CoV specific receptor. Although mRNA of ACE2 could be identified in SARS-CoV susceptible Vero E6, Huh-7, porcine POEK and PS cells, ACE2 protein expression could not be verified by several methods, suggesting that the expression level may be very low or the antibodies are not suitable for the applied experimental procedures. Different studies have revealed that recombinant ACE2 expressed on the cell surface does trigger viral permissiveness and that infection can be blocked by soluble ACE2 (Hofmann et al., 2004a; Moore et al., 2004). Additionally, ACE2 is expressed at high level in the primary target cells of SARS-CoV, namely pneumocytes and surface enterocytes of the small intestine (Hamming et al., 2004; To and Lo, 2004). Pathological findings in SARS-CoV infected patients indicated that kidney and liver are also primary targets thus underlining the results of this study identifying a kidney and a liver cell line to be susceptible to SARS-CoV. Interestingly, also cells without ACE2 expression (e.g. colonic epithelium) were found to be infected by SARS-CoV which would support the idea that accessory proteins can assist virus entry. DC-SIGN and L-SIGN which are both able to bind to the S protein of SARS-CoV were discussed, but both do not seem to be able to exclusively initiate SARS-CoV entry (Jeffers et al., 2004; Lau and Peiris, 2005)

The different infection rates in porcine POEK and PS cells may be due to a lower expression rate of the ACE2 which was shown on mRNA level. Moreover, there could be the necessity of an accessory factor for virus adsorption and entry as mentioned above. This was confirmed by others who found ACE2 expression also in cells that were not susceptible for SARS-CoV like e.g. certain endothelial cells (Chan et al., 2004; Lau and Peiris, 2005). Besides, in human and rat neural cell lines as well as in PBMCs of infected patients SARS-CoV RNA could be detected although no CPE was visible and ACE2 expression could not be determined (Ding et al., 2004; To and Lo, 2004; Wang et al., 2004; Yamashita et al., 2005).

Another possibility for lower infection rates in porcine cells may also be that the sequence homology of the human ACE2 strongly deviates from the porcine ACE2. Interestingly, infection efficiency in POEK cells could be increased by adaptation of the virus to POEK cells within 4 weeks. Consequently, future studies should clarify the question to what extent the adapted viruses differ from the original virus stock by growth comparisons and sequencing.

4.1.3 Reference gene selection for SARS-CoV quantification in infected cells

An ideal reference gene for relative quantification should be constantly transcribed in all cell types and tissues independently from internal or external influences (Radonic et al., 2004). To date, it is generally accepted that the selection of the ideal reference gene in gene expression analysis has to be done for each individual experimental setting by evaluating several genes and using the best two or three of these genes as reference. Obviously no particular gene that can be used for all experimental conditions is available. However, it is helpful to find putative candidates that can be short-listed when setting up a new experimental design.

Therefore, the expression of previously tested reference genes was determined in a setting of virus infected human cell lines (Radonic et al., 2004). Capable reference genes were evaluated using three independent methods: BestKeeper, GeNorm and the $\Delta\Delta C_T$ method, and their results were compared. Most reference genes maintained stable during SARS-CoV infection and would be useful for relative quantification. Nonetheless, summarizing the results from all five applied viruses using the three calculation tools Act, as widely used reference gene, ranked at the last position indicating that it is an unsuitable reference gene in virus infected cells. The Act gene shows significant variations with increasing degree of infection although in case of SARS-CoV infection those variations were comparably low.

Generally, only one reference gene is used by most researchers and a favourite one is certainly Act (Suzuki et al., 2000) despite its unreliability observed also by others (Selvey et al., 2001). Nonetheless, it was shown to be suitable for expression analysis using endothelial cells (Garcia-Vallejo et al., 2004) emphasizing the statement that the choice of an endogenous reference gene has to be done for each experimental setting individually.

The best genes obtained from all three tools were TBP and PPI. TBP seems to be a relative stable expressed gene during the course of virus replication of different viruses in different cells. However, as previously shown, TBP is not expressed in all tissues and thus its use may be limited (Radonic et al., 2004). PPI as a protein folding catalyst and a receptor molecule for immunosuppressive-acting drugs like cyclosporine A was shown to be a highly conserved protein among bacteria, fungi, plants and vertebrates. It is ubiquitously expressed in many tissues and thus an appropriate reference gene (Gothel and Marahiel, 1999). Interestingly, classical reference genes like β2M and GAPDH were also acceptable regarding to a stable expression in virus infected cells. All other genes showed moderate expression stability. The analysis of our data set according to the BestKeeper tool revealed very good BestKeeper indices; even Act was included into our gene panel.

These findings demonstrate the usefulness of analyzing a wide variety of reference gene candidates. However, one problem still remains to be solved; both tools, the BestKeeper and the GeNorm, can not compare paired probes. This is the great advantage of the $\Delta\Delta C_T$ method, or any other method which directly compares paired samples. From this point of view the use of a method like the $\Delta\Delta C_T$ should be applied first before considering additional tools for further elucidation of the acquired data.

In summary, TBP and PPI turned out to be the best reference genes in virus infected cells. These genes are a good point for starting reference gene selection in gene expression studies in virus infection experiments.

4.2 Characterization of novel HCoV-NL63 ORF3

The focus of the presented study was to characterize HCoV-NL63 proteins M, N and especially novel ORF3. *In silico* analysis of the ORF3 protein revealed a 225 aa (approximately 26 kDa) triple-membrane spanning protein. It could be determined that the N-terminus is facing towards the extracellular space and the 108 aa long C-terminus is at the cytosolic side. Interestingly, the HCoV-NL63 ORF3 protein, although having only 23% aa sequence homology to SARS-CoV ORF3a, shows the same topology. By analyzing deletion mutants others could verify three transmembrane domains, and they could show that the second or third transmembrane region might be responsible for Golgi localization (Tan et al., 2004b; Yuan et al., 2005). Whether all three predicted transmembrane domains exist in case of HCoV-NL63 ORF3 remains to be determined. In preliminary experiments it could be

observed that *in vitro* transcribed ORF3 is incorporated into microsomal membranes and is thus protected from proteinase K digestion. A protein band down-shift of approximately 5 kDa was detected after proteinase K digestion suggesting that not the complete C-terminus comprising 11 kDa is located in the cytosol (data not shown). This could indicate that parts of the ORF3 C-terminus are attached to the cytosolic side of the membrane as it was observed for the SARS-CoV M protein (Voss et al., 2006). Tags (N-terminal or C-terminal) had no influence on the subcellular distribution of ORF3 and viral as well as recombinant ORF3 protein is predominantly located in perinuclear regions or at the plasma membrane. Detailed localization studies revealed that ORF3 protein co-localized with an ERGIC and a Golgi compartment marker and has therefore the same localization pattern as ORF3a from SARS-CoV (Tan et al., 2004b; Yuan et al., 2005).

Moreover, the expression of viral and recombinant M and N protein in different cell lines could be verified, and it was shown that subcellular localizations are similar to other HCoVs such as SARS-CoV M and N protein (You et al., 2005; Voss et al., 2006). The M protein could be detected at the ERGIC and Golgi compartment whereas N had a more cytosolic distribution with few co-localizing areas within the ERGIC where the virus assembly and budding process occurs (Klumperman et al., 1994).

With the help of cLSM a co-localization study was performed in co-transfected HEK293T and Huh-7 cells. GFP tagged E, M and N were therefore co-transfected with a FLAG-ORF3. Additionally, the ERGIC compartment was stained to study the subcellular localization. GFP-E and M showed extensive co-localization with FLAG-ORF3 especially within the ERGIC which supports the idea that ORF3 has a similar cellular distribution as SARS-CoV ORF3a. In other experiments using a GFP-ORF3 fusion protein co-transfected with C-terminus V5-tagged E and M co-localizations could also be detected within the Golgi compartment (data not shown). For ORF3a of SARS-CoV it was already shown that it co-localizes with Golgi compartment markers and that it interacts specifically with the structural proteins E, M and S (Tan et al., 2004b; Yuan et al., 2005). The GFP-N protein is distributed in the cytosol but also shows few areas of co-localization with the ERGIC marker. As the ERGIC is the place for virus assembly and budding these co-localizations are not unexpected (Klumperman et al., 1994). Signals for N could not be detected within the nucleus or even the nucleolus like it was observed for other coronaviruses (Hiscox et al., 2001; Wurm et al., 2001). As FLAG-ORF3

co-localized with all three structural proteins within the ERGIC compartment it can be hypothesized that specific interactions of those proteins are likely.

In co-immunoprecipitation studies a confirmation of the microscopic data was performed and it could be demonstrated that GFP-E, M and N could be co-immunoprecipitated with FLAG-ORF3. As the GFP control protein also interacted with FLAG-ORF3 using different buffer conditions one cannot be certain that those interactions are specific. Others have used GFP fusion proteins and FLAG tagged proteins successfully in co-immunoprecipitation experiments confirming the results that cross-reactions are not due to tag interactions (Takuma et al., 2002). Somehow it seems that GFP alone can already interact with ORF3 in this experimental setting so that in future experiments differently tagged proteins have to be used.

Recombinant prokaryotic expressed proteins were successfully used in WB and Dot blot analysis enabling a characterization of the specific antisera to the different proteins that were also applied in IFA. After purification of recombinant proteins by affinity chromatography and subsequent WB analysis several additional bands either higher or lower than expected could be detected. Formation of multimers could lead to higher molecular weights whereas proteolytic cleavage or incomplete translation of viral proteins in *E. coli*-cells could result in lower weights.

The N protein gave additional bands at lower molecular weight which could have occurred due to incomplete translation of the relatively long protein. As an N-terminal His-tag was used for affinity purification also His tagged fragments of the N protein would be purified and eluted from the nickel-NTA column.

In case of the M protein (30 kDa including a tag) there was one additional band at approximately 45 kDa thus slightly lower than expected for a dimer. If this putative dimer formation is due to specific interactions or a result of aggregation of hydrophobic regions after heating, as described for the SARS-CoV M protein, needs to be determined (Lee et al., 2005). As eukaryotic expressed GFP-M protein was not boiled for SDS-PAGE analysis and showed no dimer formation, aggregation due to heating of prokaryotic expressed M protein seems most likely.

Interestingly, the ORF3 protein (recombinant and viral) which has a similar topology as the M protein also formed dimers and even multimers. If the only reason for ORF3 multimer

formation was aggregation of hydrophobic transmembrane domains due to heating, one would neither expect dimer formation of the hydrophilic ORF3 C-terminus protein nor of full-length eukaryotic FLAG tagged ORF3 that was sensitive to heat and thus not boiled.

In fact, it could be confirmed that recombinant ORF3 and ORF3ct protein expressed in prokaryotes, eukaryotic expressed FLAG-ORF3 and viral ORF3 all formed putative dimers and also multimers in case no reductants were added. For SARS-CoV ORF3a, which has a total of eight cysteine residues (five at the C-terminus), it was shown that disulfide bond formation between cysteine residues at cytosolic aa position 133 was responsible for dimerization. In this process two dimers seem to form homotetramers functionally working as potassium channels that modulate virus release and apparently make infected cells more permeable (Lu et al., 2006). The deduced aa sequence of HCoV-NL63 ORF3 shows four cysteine residues of which three are positioned within the C-terminus. As the full-length ORF3 proteins were found to be highly sensitive to heating and mere addition of DTT showed no effect on dimer formation, the SARS-CoV ORF3a results could initially not be verified. Instead, ORF3ct protein was used and dimer and multimer formation could be avoided after addition of DTT indicating that the C-terminus alone can dimerize with the help of disulfide bonds. This seems to be surprising, taking into account that the C-terminus is predicted to be located in the cytosol, which usually has a reducing environment inhibiting disulfide bond formation. Nonetheless, several proteins (e.g. chaperones) are activated by the formation of disulfide bridges in cytosol upon exposure to oxidative stress (Aslund and Beckwith, 1999; Jakob et al., 1999). In case of another coronavirus, TGEV, it was shown that programmed cell death was induced by a caspase-dependent pathway and that the thiol agent pyrrolidine dithiocarbamate inhibited apoptosis, suggesting that TGEV infection may lead to apoptosis via cellular oxidative stress (Eleouet et al., 1998). Thus one could hypothesize that a channel forming protein like SARS-CoV ORF3a is activated upon oxidative stress induced by viral replication. This would ensure that cell permeabilization is initiated with delay providing enough time for proper viral propagation in cells. Interestingly, other viruses like influenza A also use a disulfide linked homotetrameric protein (M_2) which has also been shown to function as an ion channel in plasma membranes of infected cells and is moreover a component of the virion envelope (Jackson et al., 1991; Hughey et al., 1992).

Another possibility for C-terminal disulfide bond formation would be that those cysteine containing parts of the C-terminus are not directly exposed to the cytosol but are partially integrated into the membrane. This would be supported by the hydropathy plot indicating that cysteine residues at positions 131 and 137 are located in hydrophobic areas of the hydrophilic

C-terminus right in proximity to the third transmembrane domain. Dimer formation could occur by intramembrane interaction of two monomers as it was already observed for the homotetrameric CorA Mg2+ transporter protein of *Salmonella enterica* serovar Typhimurium (Warren et al., 2004). With the help of mutation analysis of cysteine residues dimer formation of the complete ORF3 needs to be evaluated in further experimental procedures.

In addition to that, the expression pattern of the viral proteins ORF3, M and N were analyzed during the course of infection. By using an RT-PCR for the detection of sg mRNAs a correlation of viral replication with viral protein expression could be shown. As observed for other coronaviruses, sg mRNAs for the N protein could be detected very early post infection. A general increase of genomic and sg mRNAs could be detected 18 h post infection. Although detection of viral proteins was not possible at that point, one could suggest that this is the time when the first viral particles start to form and are released to the extracellular space at low levels. This observation is in agreement with other HCoV-NL63 infection studies in which the number of genomes increased from days two to three post inoculation and reached a plateau on day four (Schildgen et al., 2006). Ten times less MOI was used in comparison to those studies which could explain why expression levels of viral proteins are relatively low at the beginning. Additionally, the presented results emphasize that molar ratios of the different sg mRNA are roughly constant and that sg mRNAs can be detected early post infection (Stern and Kennedy, 1980; Leibowitz et al., 1981).

The N protein is the first structural viral protein to be expressed two dpi. Interestingly, it could be verified that also the viral N protein of HCoV-NL63 is able to form dimers five dpi, which was also shown for SARS-CoV N protein (He et al., 2004a; Chang et al., 2006). Dimer formation occurred by interaction of C-terminal regions of the SARS-CoV N protein and might be a conserved feature among coronaviruses possibly important for forming the ribonucleocapsid complex (Yu et al., 2005).

The expression of viral ORF3 and M is characterized by the occurrence of several protein bands. Truncated and O-glycosylated forms of SARS-CoV ORF3a have been found by others (Shen et al., 2005; Oostra et al., 2006). In the case of SARS-CoV the M protein is known to be N-glycosylated (Voss et al., 2006) whereas closely related other group II coronavirus M proteins are O-glycosylated (Rottier, 1995). It was shown that both ORF3 and M protein of group I coronavirus NL63 are N-glycosylated. By using *in silico* data putative N-glycosylation sites were identified at aa positions 3, 19 and 188 for M and 16, 119 and 126 for ORF3, respectively. For ORF3 protein only the first site is likely to be glycosylated as the

other sites are at the cytosolic side. This is emphasized by the occurrence of one additional band. For the M protein, however, all three sites are putative targets although the third site at position 188 could only be used if the M protein actually spans the membrane four times which has not been verified by experiments yet. In fact, only two additional bands were detected in case of the M protein. One reason could be that only a small proportion of M proteins have an extracellular C-terminus. This was already obverved for the related coronavirus TGEV which was shown to have only one third of its M proteins with a so-called N_{exo}-C_{exo} topology. Another reason could be that the glycosylation site at position 19 is already incorporated into the membrane and thus not accessible for glycosylation. In order to identify the exact positions a mutation analysis of those sites will have to be performed. Up to now one can only speculate why ORF3 and M are glycosylated but as studies on the M protein of MHV revealed that glycosylation has an influence on viral replication in the livers of infected mice it is possible that glycosylation plays an important role in virus host interactions (de Haan et al., 2003). Generally, glycosylation plays major roles in folding, stability and intracellular sorting of proteins as well as the generation of immune responses (Drickamer and Taylor, 1998; Van den Steen et al., 1998; Helenius and Aebi, 2001). As the ORF3a protein of SARS-CoV was already shown to be transported to the plasma membrane and is moreover able to induce antibody responses in patients, an assisting or triggering effect of the glycosylation can be assumed.

Conclusively, the ORF3 of HCoV-NL63 shows many similarities to the homologous ORF3a of SARS-CoV despite the little sequence homology of only 23% on aa level. Although an essential *in vivo* function of ORF3 proteins has yet to be elucidated, their conserved appearance in most coronaviruses and their striking similarity to the M protein lead to the assumption that ORF3 proteins may play a more important role than has been estimated so far.