

NUCLEOCAPSID PROTEIN EXPRESSION FACILITATES CORONAVIRUS REPLICATION

Barbara Schelle, Nadja Karl, Burkhard Ludewig, Stuart G. Siddell,
and Volker Thiel*

1. INTRODUCTION

The coronavirus nucleocapsid (N) protein has been implicated in a number of functions. As a structural protein, it forms a ribonucleoprotein complex with genomic RNA. However, it has also been described as an RNA-binding protein¹ that might be involved in coronavirus RNA synthesis.² Here, we used a reverse genetic approach to elucidate N protein function(s) in coronavirus replication and transcription. We could show that human coronavirus 229E (HCoV-229E) vector RNAs are greatly impaired in their ability to replicate if they lack the N gene. In contrast, efficient replication was observed if vector RNAs express N protein. Noteworthy, transcription of subgenomic mRNAs was readily detectable, irrespective of the presence or absence of N protein. Finally, by modifying the transcription signal required for the synthesis of N protein mRNA in the HCoV-229E genome, we could demonstrate selective replication of genomes that are able to express the N protein. Therefore, we conclude that the coronavirus N protein is involved in genome replication.

2. RESULTS

2.1. Analysis of Vector RNA Replication and Transcription

To study the role of N protein in coronavirus replication and transcription, we made use of our reverse genetic system for HCoV-229E.³ We produced two vector RNAs, HCoV-vec-1 and HCoV-vec-GN, that both contain the replicase gene and the gene encoding for the green fluorescent protein (GFP). HCoV-vec-GN encodes in addition the N protein, whereas HCoV-vec-1 is lacking the N gene (Figure 1a). We previously

* Barbara Schelle and Nadja Karl, University of Würzburg, Würzburg, Germany. B. Ludewig and V. Thiel, Kanton Hospital St. Gallen, 9007 St. Gallen, Switzerland. S.G. Siddell, University of Bristol, Bristol, United Kingdom.

showed that HCoV-vec-1 RNA mediates the transcription of a subgenomic mRNA, encoding GFP.⁴ Thus, we concluded, that the replicase gene products suffice for coronavirus transcription. We also observed that only a small number of green fluorescent cells (< 0.1%) can be detected upon electroporation of HCoV-vec-1 RNA into BHK-21 cells. However, if a synthetic mRNA encoding the HCoV-229E N protein was co-electroporated into BHK-21 cells, the number of green fluorescent cells significantly increased (3.4%, Figure 1b). When we electroporated HCoV-vec-GN RNA into BHK-21 cells, we again could observe less than 0.1% green fluorescent cells, and elevated numbers of GFP-expressing cells could only be detected if a N protein mRNA is co-electroporated (Figure 1b). Thus, we conclude that co-electroporation of mRNA encoding the HCoV-229E N protein with HCoV-229E-based vector RNAs encoding GFP increased the number of green fluorescent cells, even if the N protein was encoded by the vector RNA (e.g., HCoV-vec-GN). We interpret these results that the presence of N protein early after transfection of vector RNAs is important in our system for the formation of a functional replicase/transcriptase complex.

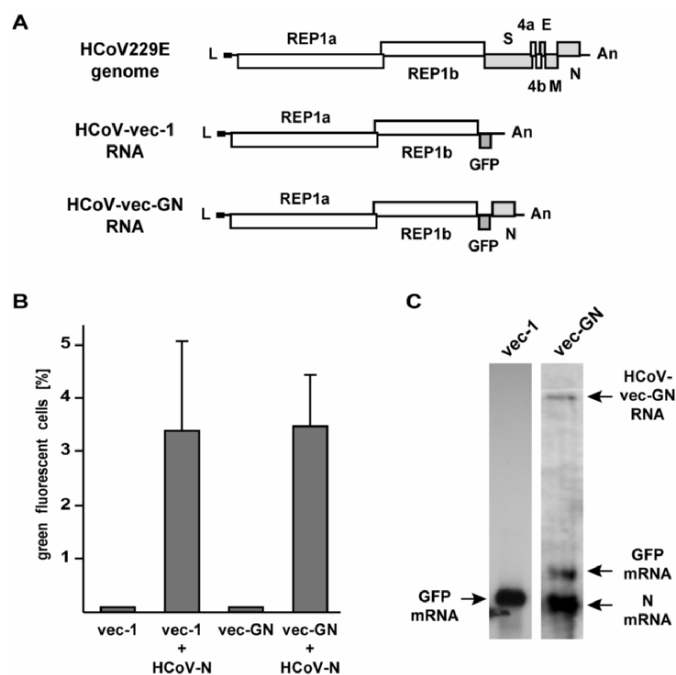


Figure 1. Transfection of vector RNAs. (A) The structural relationship of HCoV-229E and vector RNAs HCoV-vec-1 and HCoV-vec-GN are shown. Open reading frames are indicated as boxes designated by encoded gene products. L, leader RNA; An, poly(A) sequence. (B) Vector RNA HCoV-vec-1 or HCoV-vec-GN was transfected into BHK-21 cells with or without synthetic N protein mRNA as indicated. (C) Northern blot analysis of BHK-21 cells that have been transfected with N protein mRNA and vector RNAs HCoV-vec-1 or HCoV-vec-GN, respectively. Full-length vector and subgenomic mRNAs are indicated (arrows).

Although there was no difference in the number of green fluorescent cells after transfection of HCoV-vec-1 and HCoV-vec-GN RNA into BHK cells, we observed that HCoV-vec-GN-transfected cells displayed relatively intense fluorescence already 16 h post-transfection. In contrast, the intensity of green fluorescence of HCoV-vec-1-transfected cells was lower and only became apparent 48 h post-transfection. To test whether this finding was related to different replication or transcription levels of vector RNAs, we analyzed the vector-specific RNAs in transfected BHK-21 cells by Northern blot. To obtain a sufficient number of green fluorescent cells that allow for the detection of vector-specific RNAs by Northern blot analysis, we co-electroporated synthetic N protein mRNA with vector RNAs HCoV-vec-1 and HCoV-vec-GN, respectively. After co-transfection of HCoV-vec-1 RNA with N protein mRNA, we could detect a faint signal for a subgenomic mRNA encoding GFP (data not shown). To increase the number of green fluorescent cells that have been transfected with HCoV-vec-1 and N protein mRNA, we sorted green fluorescent cells prior to the isolation of poly(A)-containing RNA. With this strategy, an HCoV-vec-1 derived subgenomic mRNA encoding GFP was readily detectable, but we were unable to detect the full-length HCoV-vec-1 RNA (Figure 1c). This contrasted with the co-transfection of HCoV-vec-GN RNA and N protein mRNA, where we could easily detect (even without sorting green fluorescent cells) the full-length vector RNA and two mRNAs encoding GFP and N protein (Figure 1c). These results show that vector RNAs encoding a functional N gene were able to transcribe and replicate RNA, whereas vector RNAs lacking the N gene were able to transcribe RNA but replication of full-length vector RNA was not detectable.

Taken together, it appears that early after transfection of vector RNAs, the N protein (provided by co-transfected N protein mRNA) may be important for the establishment of a functional replicase/transcriptase complex; however, sustained N protein expression (if the N protein is expressed by the vector itself) is required for efficient replication. Accordingly, it is tempting to speculate that in natural infections, N proteins, associated with the viral genome in the ribonucleoprotein complex, may be important early in the infection to establish a functional replication/transcription complex and sustained N protein expression by the transcription of viral N protein mRNA may be required for efficient genome replication.

2.2. Selective Replication of HCoV-229E Genomes That Express N Protein

In order to test the hypothesis that N protein expression is required for efficient genome replication, we constructed recombinant HCoV-229E genomes that had been modified at the TRS of the N gene (TRS-N), a *cis*-acting RNA element that is required for the production of a subgenomic N mRNA encoding the N protein. As illustrated in Figure 2a, we modified the authentic TRS-N core sequence (UCUAAACU) to contain a stretch of three random nucleotides (UCUNNNCU). Thus, we constructed, by *in vitro* transcription, a population of 64 different full-length recombinant HCoV-229E genomes. These RNA molecules were transfected into BHK-21 cells (which are not susceptible to HCoV-229E infection), and after 3 days we isolated poly(A)-containing RNA (Figure 2b). By RT-PCR, we compared the sequence of the TRS-N region of the “input” genomes (*in vitro* transcription products) and the “re-isolated” genomes (poly(A)-containing RNA isolated 3 days post-transfection). This analysis revealed that the re-isolated genomes have clearly undergone selection, because the nucleotides at the randomized positions have shifted to a predominance of adenines (data not shown).

However, at one position within the randomized sequence, there was also a prominent uridine peak detectable. To determine the sequence of the TRS-N region on individual genomes, we cloned the RT-PCR products and determined the sequences of 44 individual clones corresponding to the input genomes and 41 individual clones corresponding to the re-isolated genomes. The result of this analysis is shown in Figure 2c. We could detect an increased number of genomes that contained the authentic (wild-type) TRS-N amongst re-isolated genomes (increase from 4.5% to 9.8%). Similarly, we observed an increased number of genomes containing the uridine within the TRS-N (NNU; increase from 20.5% to 41.5%). These two groups obviously had undergone a positive selection during amplification in BHK-21 cells. Genomes that contained one nucleotide difference compared to the authentic TRS-N or leader-TRS remained approximately at the same level (31.8% and 29.3% for input and re-isolated RNAs, respectively). Genomes that contained sequences not matching to the groups mentioned above had presumably undergone a negative selection, because their percentage dropped from 43.2% in the input genomes to 19.5% in the re-isolated genomes.

The observation that specific HCoV-229E genomes had undergone positive selection in our assay indicates that these genomes replicated preferentially. According to our hypothesis, this might be related to their ability to express N protein. Therefore, we specifically amplified, by RT-PCR, the subgenomic mRNA encoding N protein that had been produced in transfected cells. Again, the RT-PCR product was cloned, and the sequences of individual clones, corresponding to the sequences of subgenomic N protein mRNAs at the leader-body fusion sites, were determined. As expected, most N protein mRNAs (> 60%) contained either, the authentic TRS-N sequence (UCUAAACU) or the sequence of the leader TRS (UCUCAACU), confirming that these TRS elements were efficient in directing the synthesis of subgenomic mRNAs. In addition to these wild-type TRS elements, we could determine nine different N protein mRNA sequences. As shown in Figure 2d, eight of these subgenomic N protein mRNA sequences match with genomes that contain either the NNU sequence at the TRS-N (group of positively selected genomes) or only one nucleotide difference compared to the TRS-N or leader TRS (group of genomes that remained on the same level in our assay). Interestingly, only one subgenomic N protein mRNA has been detected that does not correspond to the groups of genomes that had undergone positive selection or remained at the same level during their passage in BHK-21 cells.

In summary, we could demonstrate that a population of genomes containing a stretch of randomised nucleotides within the TRS-N had undergone a selection process during the passage in BHK-21 cells. Furthermore, the vast majority of subgenomic mRNAs produced during the passage in BHK-21 cells correspond at the TRS-N region to the sequences of positively selected genomes or genomes that remained at the same level. Therefore, these data provide genetic evidence that HCoV-229E genomes that are able to express N protein confer a selective advantage for replication in nonpermissive cells.

3. CONCLUSIONS

Our data conclusively demonstrate that at least one structural protein, the N protein, is involved in coronavirus genome replication. This conclusion is based upon a reverse genetic analysis of HCoV-229E vector RNAs that showed that (i) transcription of subgenomic mRNAs can take place in the absence of N protein^{4,5} and (ii) efficient

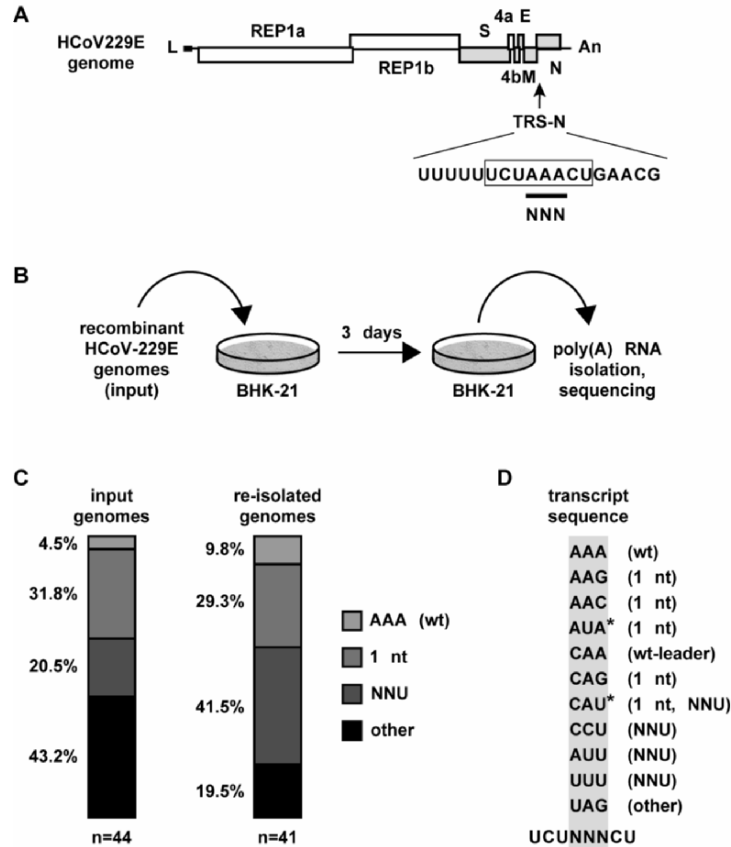


Figure 2. Selective replication of HCoV-229E genomes. (A) HCoV-229E nucleotides at the TRS-N core sequence (boxed) are shown together with the structure of the HCoV-229E genome. The authentic HCoV-229E TRS-N core sequence (UCUAAACU; randomized nucleotides [AAA] are underlined) was changed to contain a stretch of three random nucleotides (NNN). (B) Illustration of the experimental procedure. (C) Representation of genome sequences amongst input genomes and re-isolated genomes derived from 44 and 41 individual plasmid clones, respectively. The recombinant genomes were placed in four groups: group 1, recombinant genomes with the HCoV-229E wild-type sequence (AAA); group 2, recombinant genomes with a 1-nucleotide (1 nt) change compared to the TRS-N or leader TRS sequence; group 3, recombinant genomes containing a U nucleotide at the third randomized nucleotide (NNU); group 4, recombinant genomes that do not match to groups 1 to 3. The percentage of each group in the population of input and re-isolated genomes are indicated. (D) Sequences of subgenomic N protein mRNA detected in poly(A)-containing RNA 3 days post-transfection. The sequences determined at the randomized stretch of three nucleotides are shown (shaded in gray) together with corresponding groups of recombinant genomes in parentheses.

replication of vector RNAs is dependent on N protein expression. Furthermore, our conclusion is corroborated by the genetic evidence that the ability of recombinant HCoV-229E genomes to express N protein confer a selective advantage for genome replication. Thus, our data provide substantial evidence for a functional role of the N protein in coronavirus RNA synthesis. However, the nature of N protein function(s) in coronavirus RNA synthesis remains to be determined. Future studies are needed to elucidate at which

process(es) the N protein interferes in coronavirus RNA synthesis (e.g., initiation of negative-strand synthesis, recognition of TRS elements, regulation of replication/transcription levels) and how these function(s) are mediated (e.g., interaction with viral RNA, replicative proteins, or host cell proteins). It also remains to be determined whether distinct enzyme complexes, involved in replication and transcription, respectively, may exist and whether the N protein has a regulatory role in these complexes.

4. ACKNOWLEDGMENTS

This work was supported by the Swiss National Science Foundation, the Gebert-Rüf foundation, Switzerland, and the Deutsche Forschungsgemeinschaft.

5. REFERENCES

1. P. S. Masters, Localization of an RNA-binding domain in the nucleocapsid protein of the coronavirus mouse hepatitis virus, *Arch. Virol.* **125**, 141-160 (1992).
2. F. Almazan, C. Galan, and L. Enjuanes, The nucleoprotein is required for efficient coronavirus genome replication, *J. Virol.* **78**, 12683-12688 (2004).
3. V. Thiel, J. Herold, B. Schelle, and S. G. Siddell, Infectious RNA transcribed in vitro from a cDNA copy of the human coronavirus genome cloned in vaccinia virus, *J. Gen. Virol.* **82**, 1273-1281 (2001).
4. V. Thiel, J. Herold, B. Schelle, and S. G. Siddell, Viral replicase gene products suffice for coronavirus discontinuous transcription, *J. Virol.* **75**, 6676-6681 (2001).
5. V. Thiel, N. Karl, B. Schelle, P. Disterer, I. Klagge, and S. G. Siddell, Multigene RNA vector based on coronavirus transcription, *J. Virol.* **77**, 9790-9798 (2003).