Conserved RNA secondary structures and their role during discontinuous transcription in coronaviruses

Sarah Krautwurst[1], Kevin Lamkiewicz[1,2], Ramakanth Madhugiri[2,3], John Ziebuhr[2,3], Martin Hölzer[1,2] and Bashar Ibrahim[2,4]

[1] RNA Bioinformatics and High-Throughput Analysis, Friedrich Schiller University Jena, Leutragraben 1, Jena, Germany

[2] European Virus Bioinformatics Center, Friedrich Schiller University Jena, Leutragraben 1, Jena, Germany

[3] Institute of Medical Virology, Justus Liebig University Gießen, Schubertstraße 81, Gießen, Germany

[4] Chair of Bioinformatics, Matthias Schleiden Institute, Friedrich Schiller University, Jena, Germany

Background

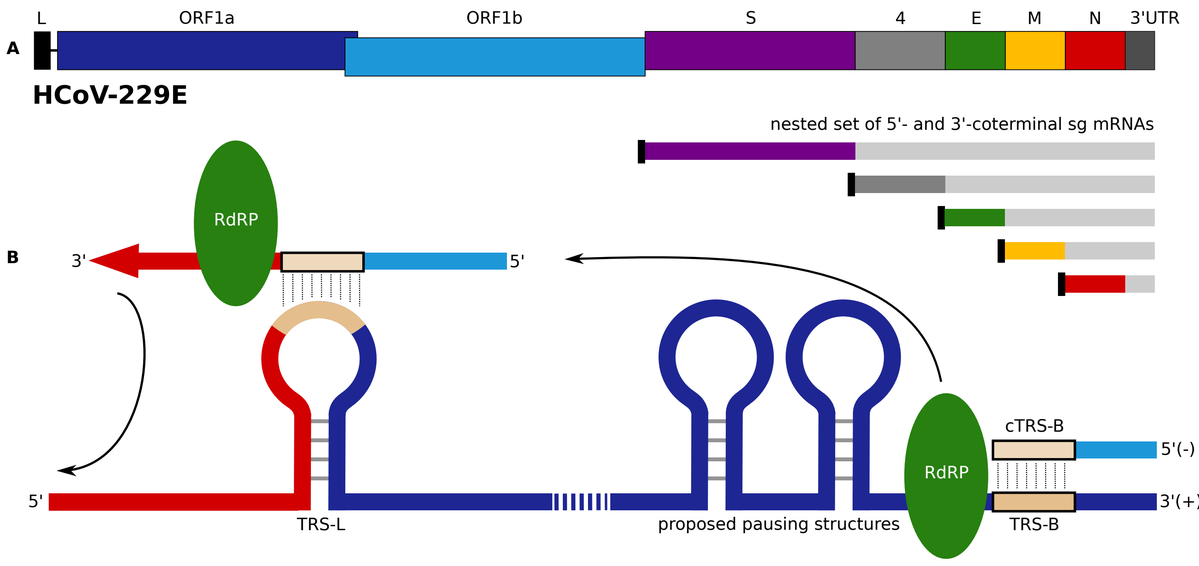


Figure 1 (A) In coronaviruses (CoV), genes coding for structural proteins are transcribed via a discontinuous mechanism by a RdRP template switch [1,2]. A body sequence (TRS-B) before each gene hybridizes with the leader sequence (TRS-L) at the 5‘-end of the genome [3,4]. (B) RdRP pausing may be facilitated by secondary structures upstream the TRS-B. (Figure adapted from [5])

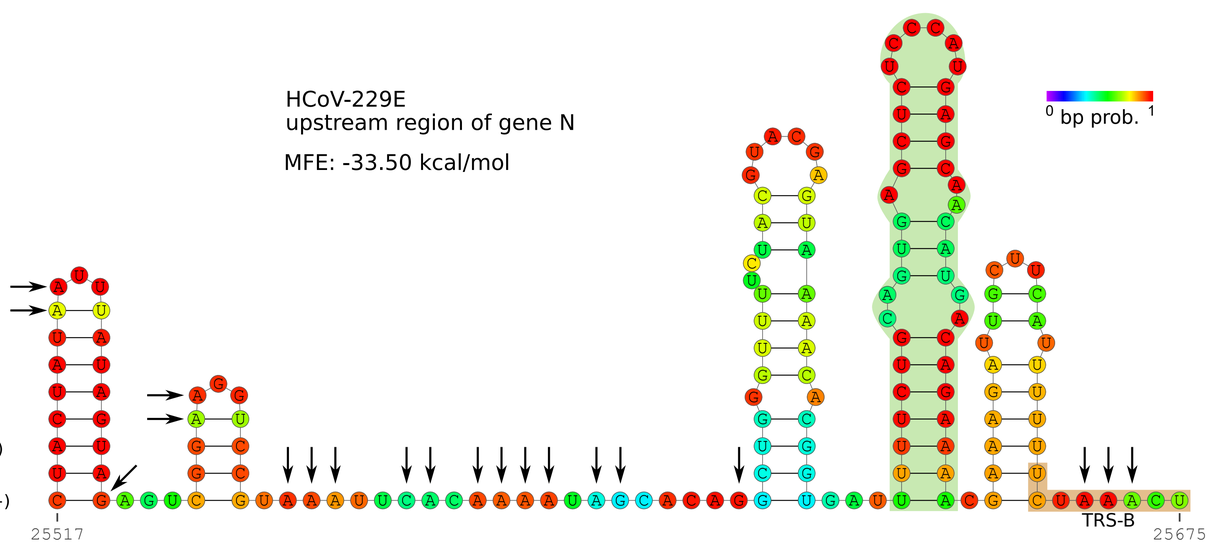


Figure 2 Initial *in silico* RNA secondary structure prediction for HCoV-229E showed multiple loops upstream the TRS-B of the N gene. Arrows indicate unpaired nucleotides as predicted by the *in vitro* DMS probing experiment, substantiating the reliability of the *in silico* fold. The highlighted structure corresponds to the one shown in Figure 4.

Results

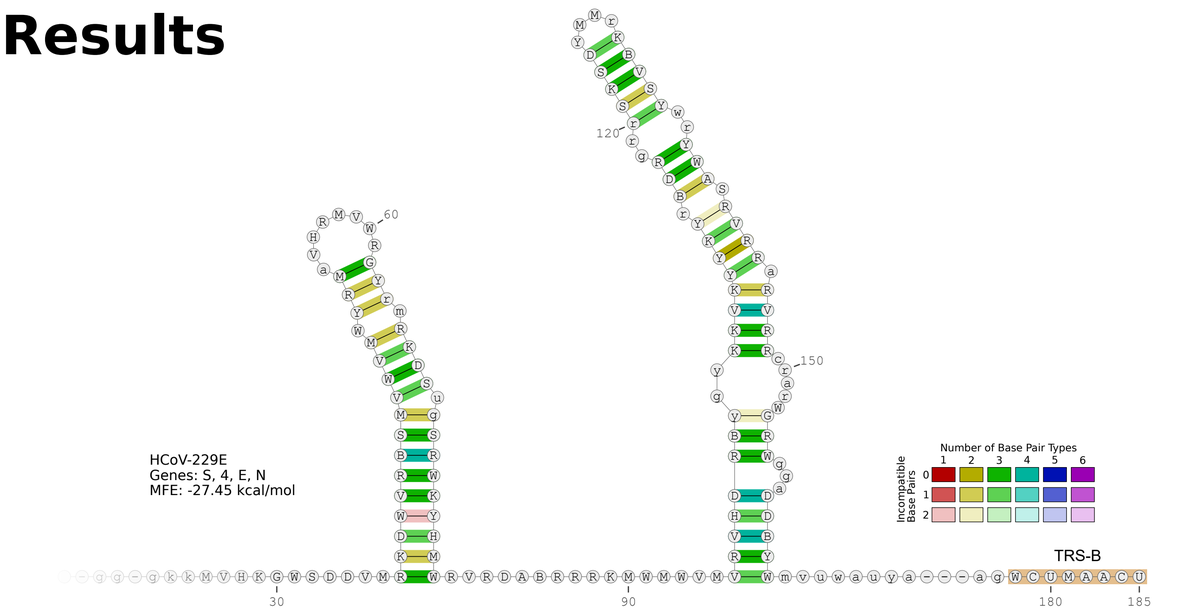


Figure 4 Alignment and consensus structure of the upstream regions of gene N for all considered *Alphacoronaviruses* reveals three distinct hairpin loops where the two most proximate to the TRS-B exhibit low conflicting nucleotides (color intensity). The highlighted structure corresponds with the individual HCoV-229E structure, supported by *in vitro* DMS probing.

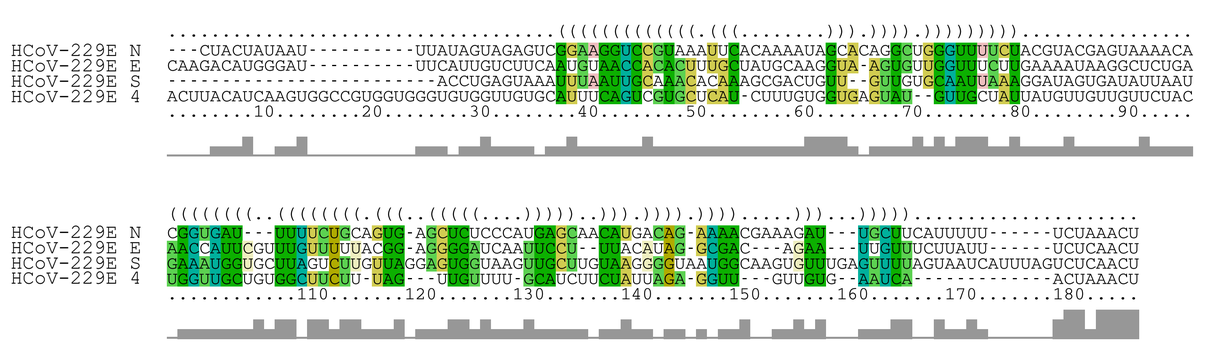


Figure 3 Alignment and consensus structure of upstream regions of all genes (exception M) of the HCoV-229E. Despite the large sequence variety the regions are conserved on structure level, indicating a potential function within the genome.

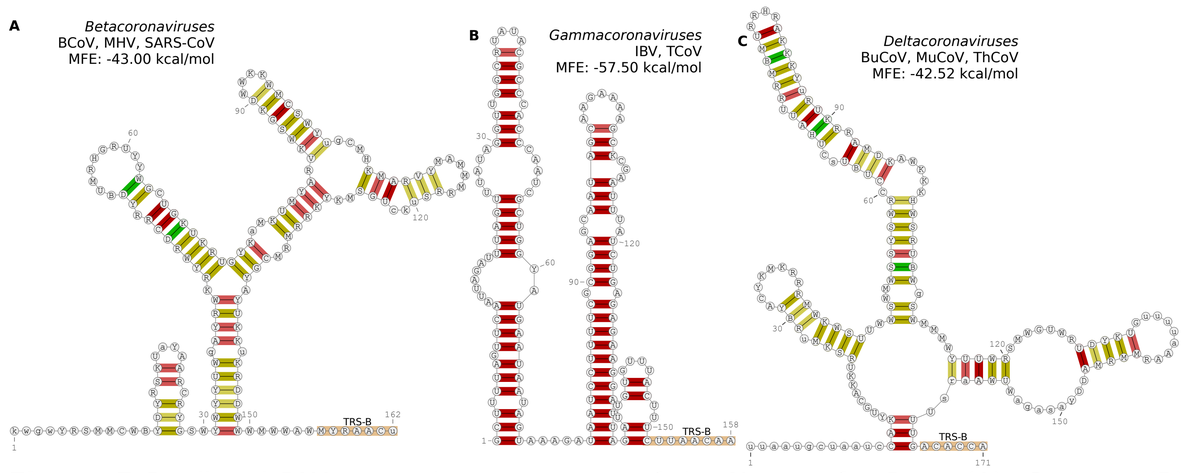


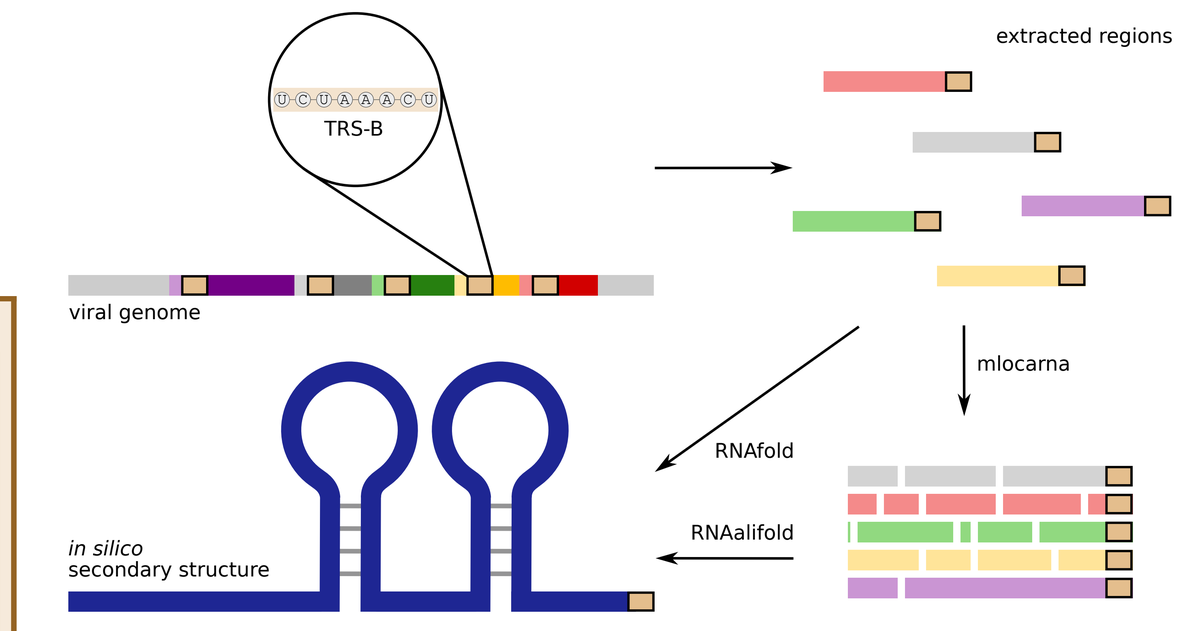
Figure 5 Predicted RNA secondary structures of the N gene for *Beta*- (A), *Gamma*- (B) and *Deltacoronaviruses* (C). For each genus, a specific and stable consensus structure can be observed. All four genera show a small distance between the TRS-B and the structure of less than 10 nucleotides which may be necessary for the RdRP to synthesize the complementary TRS-B before pausing.

Methods

After scanning the genome for the TRS-B of the respective sg mRNA, regions of 150 nt upstream were extracted. For reference genomes and manual curation of the extracted sequences, NCBI entries were used. Multiple sequence alignments were created with `mlocarna` v.1.9.2 [6]. Appropriate subsets for the alignments were chosen by applying `RNAclust` v.1.3 [7]. Secondary structure prediction was performed using `RNAalifold` v.2.4.10 [8].

Conclusions

Our findings suggest that coronaviruses harbor conserved RNA secondary structures upstream of their TRS-B sequences which are similar between species of the same genus. The structures may fulfill a task in causing or at least favoring the RdRP-pausing for the template switch during discontinuous RNA transcription. Further experimental investigations are needed to support our predictions and conclusions.



References
[1] Sawicki, S.G. et al. (1995) *Adv Exp Med Biol*, 380, 499–506.
[2] Sawicki, S.G. et al. (1998) *Adv Exp Med Biol*, 440, 215–219.
[3] Sawicki, S.G. et al. (2007) *J Virol*, 81, 20–29.
[4] Zuniga, S. et al. (2004) *J Virol*, 78, 980–994.
[5] Sola, I. et al. (2015) *Annu Rev Virol*, 2(1), 265-288.
[6] Will, S. et al. (2007) *PLOS Comput Biol*, 3, e65.
[7] RNAclust: http://www.bioinf.uni-leipzig.de/ kristin/Software/RNAclust/.
[8] Lorenz, R. et al. (2011) *Algorithms Mol Biol*, 6, 26.

Contact
sarah.krautwurst@uni-jena.de
kevin.lamkiewicz@uni-jena.de

Acknowledgements
This research was funded and supported by the BMBF (KL: 03ZZ0820A; SK: 01DL16002) and the DFG (MH: CRC 1076 A06; JZ: SFB 1021-A01 and KFO 309-P3; BI: CRC 1127 C07).

Federal Ministry of Education and Research