

## Correlated sequence signatures are present within the genomic 5'UTR RNA and NSP1 protein in coronaviruses

Piotr Sosnowski, Antonin Tidu, Eric Westhof, Franck Martin, Gilbert Eriani

## ▶ To cite this version:

Piotr Sosnowski, Antonin Tidu, Eric Westhof, Franck Martin, Gilbert Eriani. Correlated sequence signatures are present within the genomic 5'UTR RNA and NSP1 protein in coronaviruses. 27th Annual Meeting of the RNA 2022, May 2022, Boulder, United States. hal-03810281

## HAL Id: hal-03810281 https://hal.science/hal-03810281v1

Submitted on 17 Oct 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

## Correlated sequence signatures are present within the genomic 5'UTR RNA and NSP1 protein in coronaviruses



Piotr SOSNOWSKI, Antonin TIDU, Eric WESTHOF, Franck MARTIN, Gilbert ERIANI Université de Strasbourg, Institut de Biologie Moléculaire et Cellulaire

Architecture et Réactivité de l'ARN, CNRS UPR9002

2 allée Konrad Roentgen, F–67084 Strasbourg (France)



The 5'UTR part of coronavirus genomes plays key roles in the viral replication cycle and the translation of the viral mRNAs. The first 75-80 nucleotides, also called the leader sequence, are identical for the genomic mRNA and for the subgenomic mRNAs. Cooperative actions of a 5'UTR SL1 element and the nonstructural protein NSP1 are essential for both the inhibition of host mRNAs and for specific translation of viral mRNAs<sup>1-5</sup>. Sequence analyses of both the 5'UTR RNA segment and the NSP1 protein have been done for several coronaviruses with special attention to the betacoronaviruses. It appears that precise specific molecular signatures can be found in both the RNA and the NSP1 protein<sup>6</sup>. Indeed, definite sequence motifs in the RNA correlate with sequence motifs in the protein indicating a co-evolution between the 5'UTR and NSP1 in betacoronaviruses. In the Sarbecovirus subgenus, NSP1 sequences align very well. On the other hand, they do not align well with those of either the Merbecovirus or the Embecovirus<sup>6</sup>.

Here we present mutational data on 5'UTR and NSP1 from SARS-CoV-2 and SARS-CoV-1 using cell-free translation extracts. We show that some conserved key residues in the N-terminal half of the NSP1 protein are essential for evasion to the inhibitory effect of NSP1 on translation<sup>6</sup>.

<sup>1</sup> Miao Z, Tidu A, Eriani G and Martin F (2020) Secondary structure of the SARS-CoV-2 5'UTR. RNA Biology. 23, 1-10. <sup>2</sup> Tidu A, Janvier A, Schaeffer L, Sosnowski P, Kuhn L, Hammann P, Westhof E, Eriani G and Martin F (2021) The viral protein NSP1 acts as a ribosome gatekeeper for shutting down host translation and fostering SARS-CoV-2 translation. RNA, 27, 253-264

<sup>3</sup> Schubert K, Karousis ED, Jomaa A, Scaiola A, Echeverria B, Gurzeler LA, Leibundgut M, Thiel V, Mühlemann O, Ban N (2020) SARS-CoV-2 Nsp1 binds the ribosomal mRNA channel to in-hibit translation. Nat Struct Mol Biol 27, 959-966. <sup>4</sup> Thoms M, Buschauer R, Ameismeier M, Koepke L, Denk T, Hirschenberger M, Kratzat H, Hayn M, MacKens-Kiani T Cheng J, Straub JH, Stürzel CM, Fröhlich T, Berninghausen O, Becker T, Kirchhoff F, Sparrer KMJ, Beckmann R (2020) Structural basis for translational shutdown and immune evasion by the Nsp1 protein of SARS-CoV-2. Science 369, 1249-<sup>5</sup>Mendez AS, Ly M, González-Sánchez AM, Hartenian E, Ingolia NT, Cate JH, Glaunsinger BA. 2021. The N-terminal domain of SARS-CoV-2 nsp1 plays key roles in suppression of cellular gene expression and preservation of viral gene expression. Cell Rep 37,109841. <sup>6</sup> Sosnowski P, Tidu A, Ériani G, Westhof E and Martin F (2022) Correlated sequence signatures are present within the genomic 5'UTR RNA and NSP1 protein in coronaviruses. RNA 28, 729-741.







1.4<sub>1</sub>

1.2

1.0-

0.6

0.4

0.2-

0.1

⊃ 0.8-

R

C

S

ZO





EF065509 HKU5-3Cam	G––AUUUA <mark>AGAGA</mark> AUAGO	CUA <mark>GCUAU</mark> CCC	UCUCU	CUC <mark>G</mark>	UUCUCUUGC	GAACUUUGUU	UUUAAC <i>GAAC</i> UU	JAAAUAA	2006
Carboganinus									
Sarbecovirus									
Sars-CoV-2 (TYPE II)	<b>(</b> )								
MT324062	-AUUAAAGGUUU <mark>AUA</mark> -CC	UUCCCAGG-UAA	ACAAACC	AACCAAC <b>UU</b> I	JC <mark>GAUCU</mark> CUUG	UAGAUCUGUU	CUCUAAAC <mark>GAAC</mark>	UUUAAAA	2021
NC 045512	AUUAAAGGUUU <mark>AUA</mark> -CC	UUUCCCAGG-UAA	ACAA <b>A</b> CC	AACCAAC <b>UU</b> I	JC <mark>GAUCU</mark> CUUG	UAGAUCU <mark>GUU</mark>	CUCUAAAC <mark>GAAC</mark>	UUUAAAA	
мт123290.1	-UCUUUAA <mark>GGUUU</mark> A <mark>UA</mark> -CC	CU <mark>U</mark> CCC <mark>AGG-UA</mark> A	ACAA <b>A</b> CC	AACCAAC <b>UU</b> I	JC <mark>GAUCU</mark> CUUG	UAGAUCU <mark>GUU</mark>	CUCUAAAC <mark>GAAC</mark>	UUUAAAA	
мт019530.1	AUUAAA <mark>GGUUU<mark>A</mark>UA-CC</mark>	CU <mark>U</mark> CCC <mark>AGG-UA</mark> A	ACAA <b>A</b> CC	AACCAAC <b>UU</b> I	JC <mark>GAUCU</mark> CUUG	UAGAUCU <mark>GUU</mark>	CUCUAAAC <mark>GAAC</mark>	UUUAAAA	
MT020781.2	CUUAAA <mark>GGUUU<mark>AUA</mark>-CC</mark>	CU <mark>U</mark> CCC <mark>AGG-UA</mark> A	ACAA <b>A</b> CC	AACCAAC <b>UU</b> I	JC <mark>GAUCU</mark> CUUG	UAGAUCU <mark>GUU</mark>	CUCUAAAC <mark>GAAC</mark>	UUUAAAA	
MT184908.1	AUUAAAGGUUU <mark>AUA</mark> -CC	UUUCCCAGG-UAA	ACAA <b>A</b> CC	AACCAAC <b>UU</b> I	JC <mark>GAUCU</mark> CUUG	UAGAUCUGUU	CUCUAAAC <mark>GAAC</mark>	UUUAAAA	
MT152824.1	UAAAGGUUU <mark>AUA</mark> -CC	UUUCCCAGG-UAA	ACAA <b>A</b> CC	AACCAAC <b>UU</b> I	JC <mark>GAUCU</mark> CUUG	UAGAUCUGUU	CUCUAAAC <mark>GAAC</mark>	UUUAAAA	
МТ123291.2	AAAGGUUU <mark>AUA</mark> -CC	CU <mark>U</mark> CCC <mark>AGG-UA</mark> A	ACAA <b>A</b> CC	AACCAAC <b>UU</b> I	JC <mark>GAUCU</mark> CUUG	UAGAUCU <mark>GUU</mark>	CUCUAAAC <mark>GAAC</mark>	UUUAAAA	
MT163718.1	GUU <mark>GGUUU</mark> A <mark>UA</mark> -CC	CU <mark>U</mark> CSC <mark>AGG-UA</mark> A	ACAA <b>A</b> CC	AACCAAC <b>UU</b> I	JC <mark>GAUCU</mark> CUUG	UAGAUCU <mark>GUU</mark>	CUCUAAAC <mark>GAAC</mark>	UUUAAAA	
LC528233.1	GCUGUU- <mark>GUUU</mark> A <mark>UA</mark> -CC	CU <mark>U</mark> CCC <mark>AGG</mark> -UAA	ACAA <mark>A</mark> CC	AACCAAC <b>UU</b> U	JC <mark>GAUCU</mark> CUUG	UAGAUC <mark>UGUU</mark>	CUCUAAAC <mark>GAAC</mark>	UUUAAAA	2020
Sars-CoV-1 (TYPE I)									
MK211377 YN2018C Bat	LUUUAGGUUU– <mark>UU</mark> ACC	CUACCCAGG-AA-	AA <b>G</b> CC	AACCAAC-C	JC <mark>GAUCU</mark> CUUG	UAGAUCU <mark>GUU</mark>	CUCUAAAC <mark>GAAC</mark>	UUUAAAA	2018
MK211378 YN2018D Bat	UAUUAGGUUU <mark>UCU</mark> ACC	CUACCCAGG-AA-	AA <b>G</b> CC	AACCAAC-C	JC <mark>GAUCU</mark> CUUG	UAGAUCUGUU	CUCUAAAC <mark>GAAC</mark>	UUUAAAA	2018
KY770858 Anlong103 H	Bat $GUUU A - C$	CUACCA-GG-AA-		AACCAAC-C	JCGAUCUCUUG	UAGAUCUGUU	CUCUAAAC GAAC	UUUAAAA	2017
KY770859 Anlong-1121	Bat $U - UUA - C$	UACCCAGG-AA-		AACCAAC-C	JCGAUCUCUUG	UAGAUCUGUU	CUCUAAAC GAAC	UUUAAAA	2017
KY417143 Rs4081 Bat	AUAUUAGGUUU-UUACC	UACCCAGG-AA-	AA <b>G</b> CC	AACCAAC-C	JCGAUCUCUUG	UAGAUCUGUU	CUCUAAAC <mark>GAAC</mark>	UUUAAAA	2016
KY417146 Rs4231 Bat	AUAUUA <mark>GGUUU</mark> – <mark>UU</mark> ACO	CU <mark>A</mark> CCC <mark>AGG-</mark> AA-	AA <b>G</b> CC	AACCAAC <b>-C</b> I	JC <mark>GAUCU</mark> CUUG	UAGAUCU <mark>GUU</mark>	CUCUAAAC <mark>GAAC</mark>	UUUAAAA	2016
KY417150 Rs4874 Bat	AUAUUAGGUUU- <mark>UU</mark> ACC	CUACCCAGG-AA-	AA <b>G</b> CC	AACCAAC-CI	JC <mark>GAUCU</mark> CUUG	UAGAUCU <mark>GUU</mark>	CUCUAAAC <mark>GAAC</mark>	UUUAAAA	2016
DQ071615 Rp3	AUAUUAGGUUU- <mark>UU</mark> ACC	CUACCCAGG-AA-	AA <b>G</b> CC	AACCAAC-CI	JC <mark>GAUCU</mark> CUUG	UAGAUCU <mark>GUU</mark>	CUCUAAAC <mark>GAAC</mark>	UUUAAAA	2005
AY291315 Frankfurt1	AUAUUAGGUUU- <mark>UU</mark> ACC	CUACCCAGG-AA-	AA <b>G</b> CC	AACCAAC-CI	JC <mark>GAUCU</mark> CUUG	UAGAUCU <mark>GUU</mark>	CUCUAAAC <mark>GAAC</mark>	UUUAAAA	2003
NC 004718.3/Tor2	AUAUUAGGUUU- <mark>UU</mark> ACC	CUACCCAGG-AA-	AA <b>G</b> CC	AACCAAC-CI	JC <mark>GAUCU</mark> CUUG	UAGAUCU <mark>GUU</mark>	CUCUAAAC <mark>GAAC</mark>	UUUAAAA	2003
NC 004718 Sarscov	AUAUUAGGUUU- <mark>U</mark> UACC	CUACCCAGG-AA-	AA <b>G</b> CC	AACCAAC-C	JC <mark>GAUCU</mark> CUUG	UAGAUCUGUU	CUCUAAAC <mark>GAAC</mark>	UUUAAAA	2003
AY278741 Sarscov	AUAUUAGGUUU- <mark>UU</mark> ACC	CUACCCAGG-AA-	AA <b>G</b> CC	AACCAAC-C	JC <mark>GAUCU</mark> CUUG	UAGAUCU <mark>GUU</mark>	CUCUAAAC <mark>GAAC</mark>	UUUAAAA	2003
AY502924 TW11	AUAUUAGGUUU- <mark>UU</mark> ACC	CUACCCAGG-AA-	AA <b>G</b> CC	AACCAAC-C	JC <mark>GAUCU</mark> CUUG	UAGAUCU <mark>GUU</mark>	CUCUAAAC <mark>GAAC</mark>	UUUAAAA	2003
AY313906 GD69	AUAUUAGGUUU- <mark>UU</mark> ACO	CUACCCAGG-AA-	AAGCC	AACCAAC-C	JCGAUCUCUUG	UAGAUCUGUU	CUCUAAACGAAC	UUUAAAA	2003
AY304486 SZ3 Civet		CUACCCAGG-AA-	AA <b>G</b> CC	AACCAAC-C	JC <mark>GAUCU</mark> CUUG	UAGAUCUGUU	CUCUAAAC <mark>GAAC</mark>	UUUAAAA	2003

**Hibecovirus** subgenus contains a single sequence and its SL1 apical loop has similarities (UGC) with that of the **Nobecoviruses**. These two subgenera have been observed only in bats. The **Embecovirus** subgenus displays a

**Nobecoviruses**. These two subgenera have been observed only in bats. The **Embecovirus** subgenus displays a larger variation in sequence types (first sequence dates to 1997). Two subgroups exist depending on the apical loop of SL1, either UGC (like in the Nobe- and Hibecovirus) or CGU. The sequences of the **Merbecovirus** subgenus display a much larger conservation with a 3-nt loop, often -CUA. Two subgroups can be distinguished in the **Sarbecovirus** subgenus: the SARS-CoV-1 and SARS-CoV-2. The sequence conservation is high and for SARS-CoV-1, it extends over 15 years. The similarities between the two subgroups are extensive: SL2 and SL3 are identical, the stem of SL1 presents only a change from a UoG to a U–A pair. The main differences are in SL1B, the first nucleotide of the SL1 hairpin loop (A>U), two positions 3' of SL2 following SL1, after the conserved stretch AACCAAC, there is a conserved UU in SARS-CoV-2 that is replaced by a single C in SARS-CoV-1. a single C in SARS- CoV-1.



Amino acid residues specific of SARS-CoV-1 that were swapped into a SARS-CoV-2 background. They were chosen on the 'front' side of NSP1 where is found the previously studied R99A mutation (Mendez et al. 2021). R99A abrogates viral translation evasion while reducing the translation inhibition activity of NSP1 against cellular mRNA. This suggested that SL1 might interact by contacts with residues located on the same front side.

Ten NSP1 proteins were overproduced and purified, corresponding to 11 amino acid residues located on the 'front' side.





SARS-CoV NSP1 inhibits host mRNA translation by binding to the 40S ribosomal subunit and blocking the mRNA channel while promoting translation of viral mRNA. We show here that mutations in the N-terminal domain of NSP1 can induce distinct types of effects on mRNA translation. Several mutations such as the deletion of the first residues of the N-terminal end or E102I mutant abrogates escape of SARS-CoV-2 leader mRNA. A partial loss of inhibitory capacity was observed for several mutants suggesting decreased stability on the ribosome. We also show that the presence of an A in the apical loop of SL1 confers the capacity for recognition by NSP1 originating from SARS-CoV-1 and that identity swapping can be achieved by substitution of this residue. Finally, residues specific to SARS-CoV-1 or -2 have been identified allowing recognition switches of the corresponding SL1. Altogether, data show that both the 5'UTR and the NSP1 protein are linked functionally and therefore are forced to coevolve to interact cooperatively with highly conserved elements of the ribosomal machinery and initiation cofactor proteins.

The level of translation inhibition of SARS-CoV-1 NSP1 is significantly lower than that with SARS-CoV-2 NSP1. Similarly, the translation evasion of cognate viral mRNA constructs was also less efficient with SARS-CoV-1 NSP1 than with SARS-CoV-2 NSP1.

Agence Nationale pour la Recherche (ANR-17-CE12-0025-01, ANR-17-CE11-0024, ANR-20-COVI- 0078), Fondation pour la Recherche Médicale (project CoronalRES), Fondation Bettencourt Schueller, University of Strasbourg and the Centre National de la Recherche Scientifique.