



How the Replication and Transcription Complex Functions in Jumping Transcription of SARS-CoV-2

Jianguang Liang^{1†}, Jinsong Shi^{2†}, Shunmei Chen³, Guangyou Duan⁴, Fan Yang², Zhi Cheng⁵, Xin Li⁵, Jishou Ruan⁶, Dong Mi^{7*} and Shan Gao^{5*}

¹School of Pharmacy, Changzhou University, Changzhou, China, ²National Clinical Research Center of Kidney Disease, Jinling Hospital, Nanjing University School of Medicine, Nanjing, China, ³Yunnan Key Laboratory of Stem Cell and Regenerative Medicine, Biomedical Engineering Research Center, Kunming Medical University, Kunming, China, ⁴School of Life Sciences, Qilu Normal University, Jinan, China, ⁵College of Life Sciences, Nankai University, Tianjin, China, ⁶School of Mathematical Sciences, Nankai University, Tianjin, China, ⁷Department of Clinical Laboratory, Affiliated Maternity Hospital, Nankai University, Tianjin, China

OPEN ACCESS

Edited by:

Tao Huang,

Shanghai Institute of Nutrition and Health (CAS), China

Reviewed by:

Xue Yao,

Harvard University, United States

Yanna Shen,

Tianjin Medical University, China

*Correspondence:

Dong Mi

mimidong@163.com

Shan Gao

gao_shan@mail.nankai.edu.cn

[†]These authors have contributed equally to this work

Specialty section:

This article was submitted to Computational Genomics,

a section of the journal

Frontiers in Genetics

Received: 25 March 2022

Accepted: 13 May 2022

Published: 30 May 2022

Citation:

Liang J, Shi J, Chen S, Duan G, Yang F, Cheng Z, Li X, Ruan J, Mi D and Gao S (2022) How the Replication and Transcription Complex Functions in Jumping Transcription of SARS-CoV-2.

Front. Genet. 13:904513.

doi: 10.3389/fgene.2022.904513

Background: Coronavirus disease 2019 (COVID-19) is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Although unprecedented efforts are underway to develop therapeutic strategies against this disease, scientists have acquired only a little knowledge regarding the structures and functions of the CoV replication and transcription complex (RTC). Ascertaining all the RTC components and the arrangement of them is an indispensably step for the eventual determination of its global structure, leading to completely understanding all of its functions at the molecular level.

Results: The main results include: 1) hairpins containing the canonical and non-canonical NSP15 cleavage motifs are canonical and non-canonical transcription regulatory sequence (TRS) hairpins; 2) TRS hairpins can be used to identify recombination regions in CoV genomes; 3) RNA methylation participates in the determination of the local RNA structures in CoVs by affecting the formation of base pairing; and 4) The eventual determination of the CoV RTC global structure needs to consider METTL3 in the experimental design.

Conclusions: In the present study, we proposed the theoretical arrangement of NSP12-15 and METTL3 in the global RTC structure and constructed a model to answer how the RTC functions in the jumping transcription of CoVs. As the most important finding, TRS hairpins were reported for the first time to interpret NSP15 cleavage, RNA methylation of CoVs and their association at the molecular level. Our findings enrich fundamental knowledge in the field of gene expression and its regulation, providing a crucial basis for future studies.

Keywords: coronavirus, RNA methylation, nanopore, TRS hairpin, METTL3

INTRODUCTION

Coronavirus disease 2019 (COVID-19) is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (Li et al., 2020) (Duan et al., 2020) with a genome of ~30 kb (Jiayuan et al., 2020). By reanalyzing public data (Kim et al., 2020a), we determined that a SARS-CoV-2 genome has 12 genes, which are *spike* (*S*), *envelope* (*E*), *membrane* (*M*), *nucleocapsid* (*N*), and *ORF1a*, *1b*, *3a*, *6*, *7a*, *7b*, *8* and *10* (Li et al., 2021a). The *ORF1a* and *1b* genes encode 16 non-structural proteins (NSPs), named NSP1 through NSP16 (Silva et al., 2020), while the other 10 genes encode four structural proteins (*S*, *E*, *M* and *N*) and six accessory proteins (*ORF3a*, *6*, *7a*, *7b*, *8* and *10*). Among the above 26 proteins, NSP4-16 are highly conserved in all known CoVs and have been experimentally demonstrated or predicted to be critical enzymes in CoV RNA synthesis and modification (Denison et al., 2011), particularly including: NSP12, RNA-dependent RNA polymerase (RdRP) (Yan et al., 2020); NSP13, RNA helicase-ATPase (*Hel*); NSP14, RNA exoribonuclease (*ExoN*) and N7 methyltransferase (*MTase*); NSP15 endoribonuclease (*EndoU*) (Kim et al., 2020b); and NSP16, RNA 2'-O-MTase.

NSP1-16 assemble into a replication and transcription complex (RTC) (Yan et al., 2020). The basic function of the RTC is RNA synthesis: it synthesizes genomic RNAs (gRNAs) for replication or transcription of the *ORF1a*, *1b* genes, while it synthesizes subgenomic RNAs (sgRNAs) for jumping transcription of the other 10 genes (Kim et al., 2020a). In 1998, the “leader-to-body fusion” model (Sawicki et al., 1998) was proposed to explain the jumping transcription, however, the molecular basis of this model was unknown until our previous study in 2020 (Li et al., 2021a). In our previous study (Li et al., 2021a), we provided a molecular basis for the “leader-to-body fusion” model by identifying the cleavage sites of NSP15 and proposed a negative feedback model to explain the regulation of CoV replication and transcription. In addition, we revealed that the jumping transcription and recombination of CoVs share the same molecular mechanism (Li et al., 2021a), which causes rapid mutation and inevitably outbreaks of CoVs. These findings are vital for the further investigation of CoV transcription and recombination. However, there will be a long way to completely understand how the RTC functions in the jumping transcription at the molecular level.

For a complete understanding of CoV replication and transcription, particularly the jumping transcription, much research (Yan et al., 2020) (Kim et al., 2020b) (Hillen et al., 2020) has been conducted to determine the global structure of the SARS-CoV-2 RTC, since the outbreak of SARS-CoV-2 in 2019. Although some single protein structures (e.g., NSP15 (Kim et al., 2020b)) and local structures of the RTC (i.e. NSP7&8&12&13 (Yan et al., 2020) and NSP7&8&12 (Hillen et al., 2020)) have been determined, the global structure and all components of RTC are still unknown. As the global structure of the CoV RTC cannot be determined by simple use any one of current methods (i.e., X-ray, NMR and Cryo-EM), ascertaining all the RTC components and the arrangement of them is an indispensably step for the eventual determination of its global structure, leading to completely understanding all of its functions at the molecular level. In the

present study, we aimed to determine the theoretical arrangement of NSP12-16 in the global structure of the CoV RTC by comprehensive analysis of data from different sources, and to preliminarily elucidate how the RTC functions in the jumping transcription of CoVs at the molecular level.

RESULTS

Jumping Transcription, TRS and NSP15 Cleavage Site

First, we provide a brief introduction to the jumping transcription of CoVs, the “leader-to-body fusion” model proposed in an early study (Sawicki et al., 1998) and its molecular basis proposed in our recent study (Li et al., 2021a). In the “leader-to-body fusion” model, the realization of jumping transcription requires transcription regulatory sequences (TRSs), which include leader TRSs (TRS-Ls) and body transcription regulatory sequences (TRS-Bs). Each CoV genome contains a TRS-L in the 5' untranslated region (UTR) and several TRS-Bs located in the upstreams of genes except *ORF1a* and *1b*. CoV replication and transcription require gRNAs(+) as templates for the synthesis of antisense genomic RNAs [gRNAs(-)] and antisense subgenomic RNAs [sgRNAs(-)] by RdRP. When RdRP pauses, as it crosses a TRS-B and switches the template to the TRS-L, sgRNAs(-) are formed through jumping transcription (also referred to as discontinuous transcription, polymerase jumping or template switching). Otherwise, RdRP reads gRNAs(+) continuously, without interruption, resulting in gRNAs(-). Thereafter, gRNAs(-) and sgRNAs(-) are used as templates to synthesize gRNAs(+) and sgRNAs(+), respectively; gRNAs(+) and sgRNAs(+) are used as templates for the translation of NSP1-16 and the other 10 proteins (*S*, *E*, *M*, *N*, and *ORF3a*, *6*, *7a*, *7b*, *8* and *10*), respectively. In our previous study (Li et al., 2021a), we provided a molecular basis for the “leader-to-body fusion” model by identifying the reverse complimentary sequences of TRS-Bs [denoted as TRS-Bs(-)] as the NSP15 cleavage sites, which actually functions in the regulation of CoV regulation. NSP15 cleaves gRNAs(-) and sgRNAs(-) at TRS-Bs(-). Then, the free 3' ends (~6 nt) of TRS-Bs(-) hybridize TRS-Ls to realize “leader-to-body fusion”. These findings associated the investigation of TRSs to that of NSP15 cleavage sites.

In our previous study (Bei et al., 2022), we made a generalization that a TRS motif is a (6~8-nt long for CoVs) consensus sequence beginning with at least an adenosine residue (**A**), and enriched with A and followed by C, based on the analysis of 1,265 CoV genome sequences (**Materials and Methods**). We defined that the antisense sequence of a TRS motif as the motif of the corresponding NSP15 cleavage site (the NSP15 cleavage motif). For example, the canonical TRS motif of SARS-CoV-2 and the corresponding NSP15 cleavage motif are ACGAAC and GTTCGT, respectively. We defined the TRS motif in the TRS-L as the canonical TRS motif. Thus, the canonical TRS motif is unique to a CoV genome, while the TRS motifs in TRS-Bs can be canonical TRS motifs or non-canonical TRS motifs with little nucleotide (nt) differences. By these definitions, we determined canonical TRS motifs of all viruses in the order *Nidovirales*

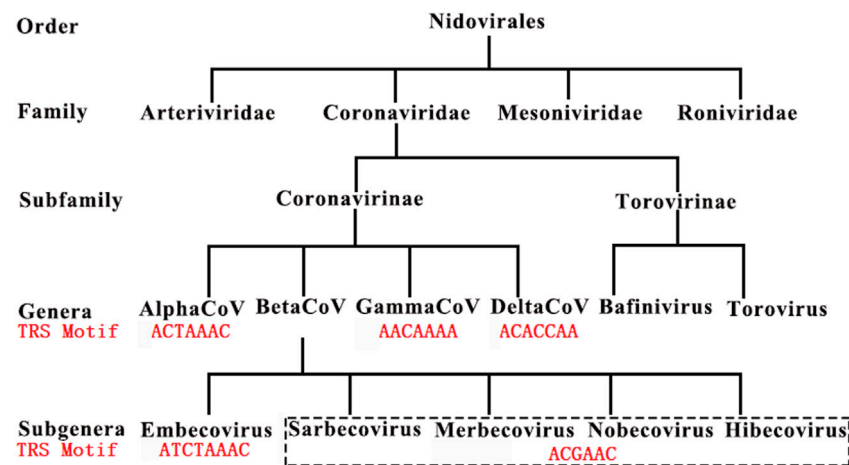


FIGURE 1 | Canonical TRS motifs in Coronaviridae. *Embecovirus*, *Sarbecovirus*, *Merbecovirus*, *Nobecovirus* and *Hibecovirus* are also defined as *Betacoronavirus* subgroups A, B, C, D and E. SARS-CoV and SARS-CoV-2 belong to *Betacoronavirus* subgroup B. These canonical TRS motifs (in red color) of viruses in Coronaviridae have been reported in our previous study (Bei et al., 2022).

(Figure 1) and corrected some canonical TRS motifs reported in the previous studies. For instance, the canonical TRS motifs of mouse hepatitis virus (MHV), transmissible gastroenteritis virus (TGEV), Canada goose coronavirus (Goose-CoV) and beluga whale coronavirus (BWCoV) were corrected from CTAAAC (Grossoehme et al., 2009), CTAAAC (Sola et al., 2005), CTTAACAAA (Papineau et al., 2019) and AAACA (Mihindukulasuriya et al., 2008) to ATCTAAAC, ACTAAAC, AACAAAA and AACAAAA, respectively. Canonical TRS motifs are highly conserved in *Alphacoronavirus*, *Gammacoronavirus*, *Deltacoronavirus* and *Betacoronavirus* genera except the subgroup A (Figure 1). *Betacoronavirus* subgroup A has the canonical TRS motif ATCTAAAC, which is different from ACGAAC in *Betacoronavirus* subgroup B, C, D and E. Different from *Betacoronavirus* subgroup B, *Betacoronavirus* subgroup A, C, D and E, *Alphacoronavirus*, *Gammacoronavirus* and *Deltacoronavirus* have non-canonical TRS motifs in the TRS-Bs of four structural genes (*S*, *E*, *M* and *N*), which were caused by mutations during evolution. These TRS motif mutations resulted in the attenuation of CoVs in *Betacoronavirus* subgroup A, D and E by down-regulating the transcription of CoV genes except *ORF1a* and *1b* (Li et al., 2021b). This confirmed that TRSs (Actually revealed as the NSP15 cleavage sites (Li et al., 2021a)) function in the regulation of CoV transcription (Yount et al., 2006). Furthermore, a previous study reported that the recognition of a TRS (Actually revealed as the NSP15 cleavage site (Li et al., 2021a)) is independent on its motif, but dependent on its context (Yount et al., 2006).

NSP15 Cleavage, RNA Methylation and TRS Hairpin

A previous study (Kim et al., 2020a) reported that RNA methylation sites containing the “AAGAA-like” motif (including AAGAA and other A/G-rich sequences) are present throughout the SARS-CoV-2 genome, particularly enriched in

genomic positions 28,500-29,500. This study used Nanopore RNA-seq (Xu et al., 2019), a direct RNA sequencing method, which can be used to measure RNA methylation at 1-nt resolution although it has a high error rate. By analyzing the Nanopore RNA-seq data, the previous study (Kim et al., 2020a) concluded that the methylated RNAs have shorter 3' polyA tails than the unmethylated ones in SARS-CoV-2. Although the type of RNA methylation was unknown, the previous study (Kim et al., 2020a) proposed that the “AAGAA-like” motif associates with the lengths of 3' polyA tails of gRNAs and sgRNAs. However, the previous study did not answer the following questions: 1) it was not explained that what functions the internal methylation sites have, as they are far from 3' ends, thus unlikely to contribute to the lengths of 3' polyA tails; and 2) the extremely high ratio between sense and antisense reads (Li et al., 2021a) may result from quick degradation of the antisense nascent RNAs due to their shorter 3' polyA tails, however, the “AAGAA-like” motif occurs in both sense and antisense strands at a similar frequency. Notably, the previous study (Kim et al., 2020a) shouldn't have neglected the analysis of the “AAGAA-like” motif on the antisense strand, since only very a few antisense reads from the Nanopore RNA-seq data were obtained for analysis. Therefore, we proposed that RNA methylation sites containing the “AAGAA-like” motif may have other biological functions and conducted further analysis.

Different from the previous study (Kim et al., 2020a), our study focused on the analysis of the “AAGAA-like” motif on the antisense strand of the SARS-CoV-2 genome, particularly the association between the “AAGAA-like” motif and the TRS or corresponding NSP15 cleavage motifs. As a result, we discovered that the “AAGAA-like” motif co-occurred with the NSP15 cleavage motif GTTCGT of four genes (*S*, *ORF6*, *7a* and *8*). In our previous study (Liu et al., 2018), complemented palindrome sequences in genomes of viruses in *Betacoronavirus* subgroup B have been investigated and most of them are semipalindromic or heteropalindromic. These complemented palindrome sequences

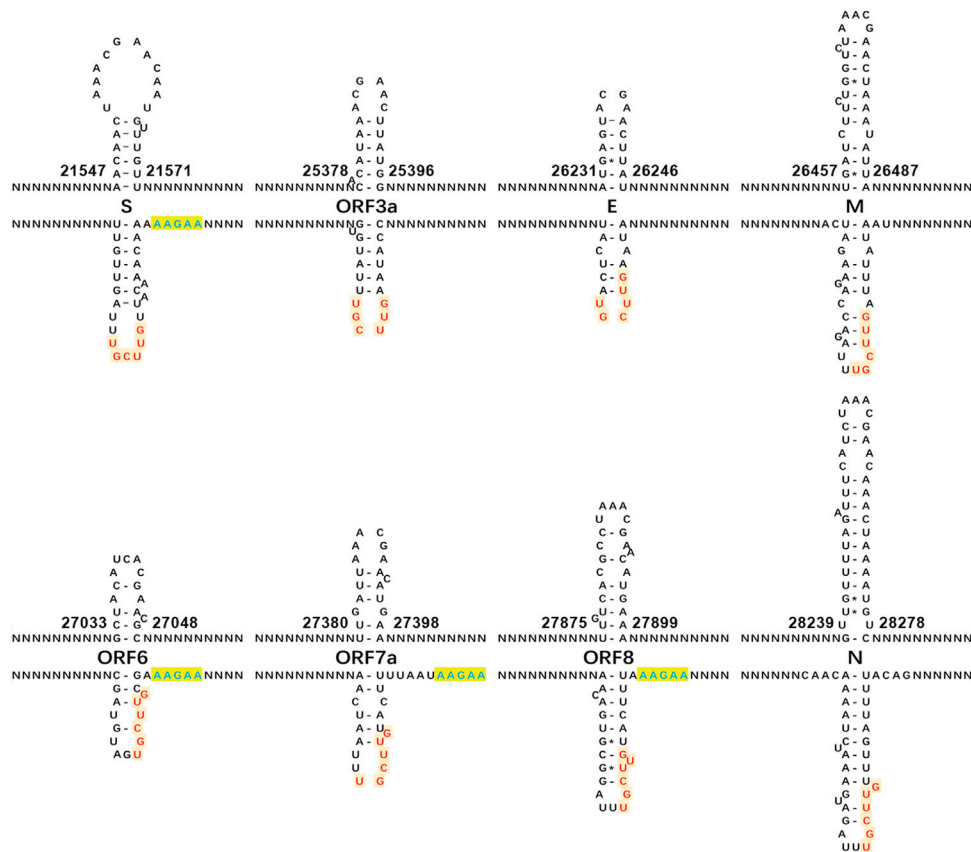


FIGURE 2 | Canonical TRS hairpins in SARS-CoV-2. The canonical transcription regulatory sequence (TRS) motif ACGAAC is present in the upstreams of eight genes (*S*, *E*, *M*, *N*, and *ORF3a*, 6, 7a and 8). Read on the antisense strands of the SARS-CoV-2 genome (GenBank: MN908947.3), “AAGAA” (in blue color) or “AAACH” (**Supplementary Table S1**) represents an RNA methylation site, while “GUUCGU” (in red color) represents a NSP15 cleavage site. The positions are the start and end positions of hairpins in the SARS-CoV-2 genome. NSP15 cleaves a single-strand RNA after U (indicated by arrows). In the present study, we defined: (1) the hairpins containing the canonical and non-canonical NSP15 cleavage sites are canonical and non-canonical TRS hairpins, respectively; and (2) the hairpins opposite to the TRS hairpins as the opposite TRS hairpins.

containing A-rich and T-rich regions form hairpins. The “AAGAA-like” and GTTCGT motifs are located in the A-rich and T-rich regions. Thus, the association between the “AAGAA-like” and GTTCGT motifs was discovered by analysis of TRS hairpins of the four genes (**Figure 2**). For analysis of TRS hairpins, we defined: 1) hairpins containing the canonical and non-canonical NSP15 cleavage sites are canonical and non-canonical TRS hairpins, respectively; and 2) hairpins opposite to TRS hairpins are opposite TRS hairpins (**Figure 2**). However, the formation of opposite TRS hairpins is uncertain, as all complemented palindrome sequences forming the TRS and opposite TRS hairpins are asymmetric (semipalindromic or heteropalindromic). Among the 10 genes, eight (*S*, *E*, *M*, *N*, *ORF1a*, *1b*, *3a*, 6, *7a*, and 8) have canonical TRS hairpins and two (*ORF7b* and 10) may have non-canonical TRS hairpins (**Supplementary Table S1**). Non-canonical TRS hairpins have been reported in seven common recombination regions in one of our previous studies (Li et al., 2021b) and identified in five recombination events (**Figure 3**) in another one of our previous studies (Li et al., 2021a). Therefore, TRS hairpins can

be used to identify recombination regions in CoV genomes. NSP15 cleaves the canonical TRS hairpins of the seven genes at canonical breakpoints, whereas it cleaves the canonical TRS hairpin of *ORF3a* at an unexpected breakpoint “GTTCGTTTAT|N” (the NSP15 cleavage motif is underlined; the vertical line indicates the breakpoint and N represents any nt), rather than the end of the canonical NSP15 cleavage motif “GTTCGT|TTATN”. According to our definitions, “GTTCGT|TTATN” and “GTTCGTTTAT|N” are canonical and non-canonical NSP15 breakpoints, respectively. The discovery of non-canonical TRS hairpins and non-canonical NSP15 breakpoints indicated that the recognition of NSP15 cleavage sites is structure-based rather than sequence-based.

How RTC Functions in Jumping Transcription

Since several A-rich and T-rich regions are alternatively present around each NSP15 cleavage site, many hypothetical TRS hairpins (**Figure 4A–C**) containing the NSP15 cleavage site

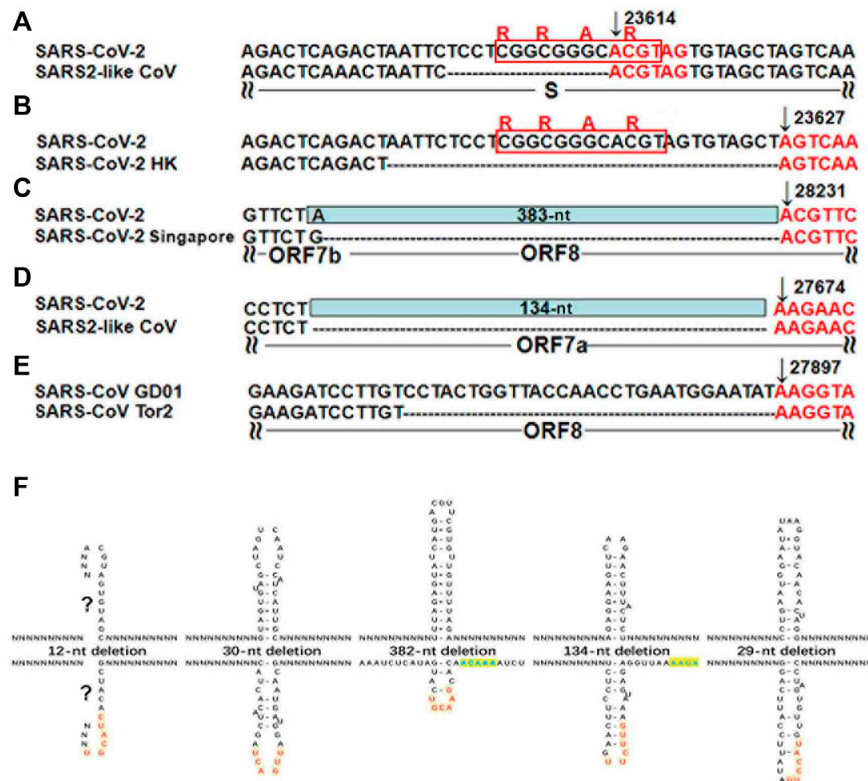
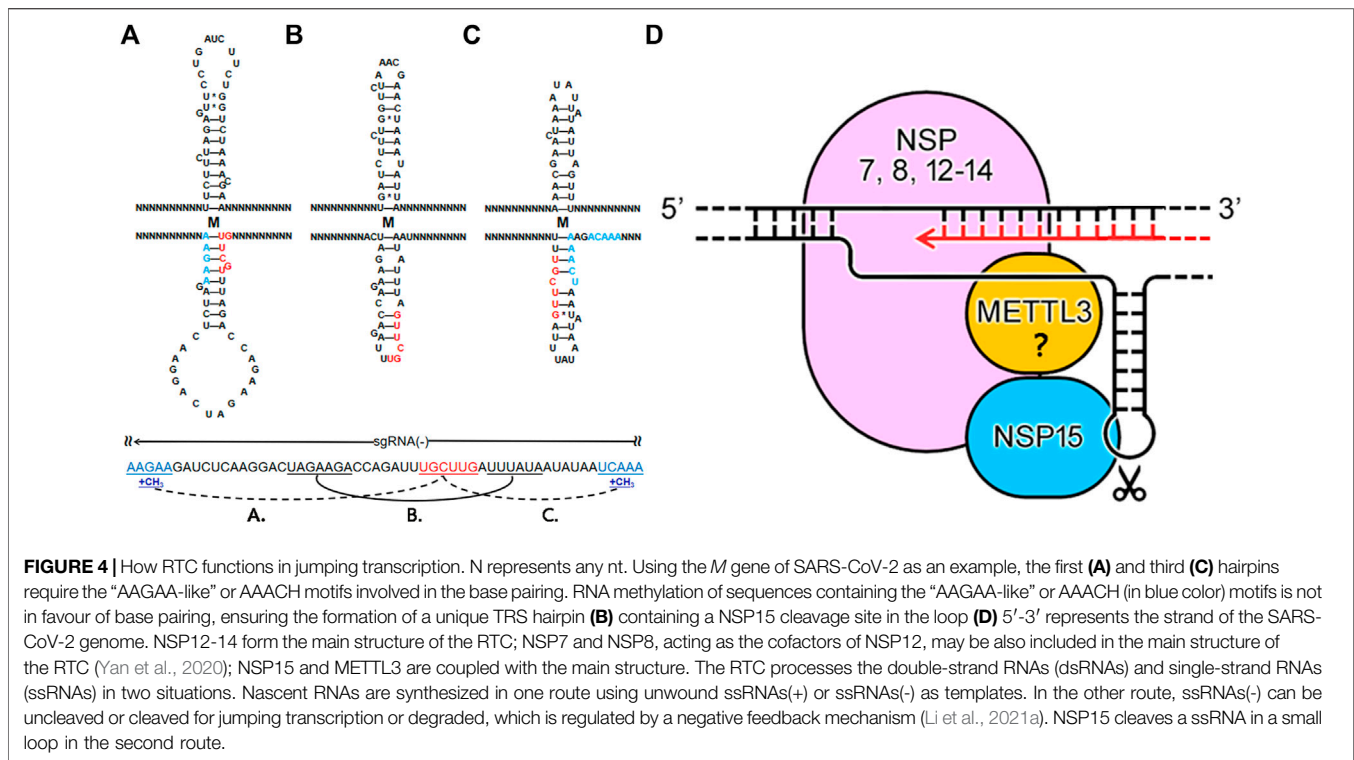


FIGURE 3 | TRS hairpins in five recombination regions. (A–E) have already been published in our previous study (Li et al., 2021a). N represents any nt. All the positions were annotated on the SARS-CoV (GenBank: AY278489) or SARS-CoV-2 (GenBank: MN908947) genomes. (A). The genome (GenBank: MN996532) of the SARS2-like CoV strain RaTG13 from bats is used to show the 12-nt deletion; (B). The genome (GenBank: EPI_ISL_417443) of the SARS-CoV-2 strain Hongkong is used to show the 30-nt deletion; (C). The genomes (GenBank: EPI_ISL_414378, EPI_ISL_414379 and EPI_ISL_414380) of three SARS-CoV-2 strains from Singapore are used to show the 382-nt deletion; (D). The genome (GenBank: MT457390) of the mink SARS2-like CoV strain is used to show the 134-nt deletion; (E). The genome (GenBank: AY274119) of the SARS-CoV strain Tor2 is used to show the 29-nt deletion. (F). These recombinant events occurred at the non-canonical NSP15 breakpoints that also end with at least an uridine residue (“U”), due to the cleavage of the non-canonical TRS hairpins.

can form. Thus, to investigate if a unique TRS hairpin can be formed, we further analyzed the association between the “AAGAA-like” and GTTCGT motifs in all possible TRS hairpins of the eight genes (Supplementary Table S1) using 1,265 CoV genome sequences (Materials and Methods), leading to discovery of the association between RNA methylation and NSP15 cleavage. Here, we illustrate how the association was discovered, using the *M* gene of SARS-CoV-2 as an example (Figure 4). The minimum free energies (MFEs) of three possible TRS hairpins in the *M* gene were estimated as -2.50, -4.00 and -4.90 kcal/mol (Materials and Methods). Although the third hairpin (Figure 4C) is the most stable one, the difference of MFEs between the second (Figure 4B) and third hairpins is marginal. The first (Figure 4A) and third hairpins require the “AAGAA-like” and AAACH (Detailed later) motifs involved in the base pairing, respectively. However, RNA methylation (e.g., m6A) of these motifs is not in favour of base pairing in the first and third hairpins. Thus, only the second hairpin was able to form. We proposed that RNA methylation participates in the determination of the local RNA structures in CoVs by affecting the formation of base pairing. RNA methylation of sequences containing the “AAGAA-like” or AAACH motifs significantly

reduces the possibility of formation of many hairpins, ensuring the formation of a unique TRS hairpin (Figure 4B) in all likelihood. In the unique TRS hairpin, the NSP15 cleavage site exposes in a small loop, which facilitates the contacts of NSP15, while the loop of the opposite TRS hairpin may not contain uridine residues for NSP15 cleavage. The structure of this small loop can be used to explain the results of mutation experiments in a previous study (Yount et al., 2006) that the recognition of a TRS (Actually revealed as the NSP15 cleavage site (Li et al., 2021a)) is independent on its motif, but dependent on its context. The TRS hairpin can be used to explain the discovery that the recognition of NSP15 cleavage sites is structure-based (TRS hairpin) rather than sequence-based (NSP15 cleavage motif). The above results indicated that TRS hairpins in nascent gRNAs(-) are indispensable for the functions of the RTC in jumping transcription (Figure 4D).

The following topic is which enzyme is responsible for the internal methylation of CoV RNAs, which is supposed to be done before the NSP15 cleavage for jumping transcription. A recent study reported that NSP14 (no structure data available) and NSP10&16 (PDB: 7BQ7), as N7 and 2'-O-MTase respectively (Introduction), are crucial for RNA cap



formation (Krafčikova et al., 2020). This suggested that NSP14 and NSP10&16 are unlikely to function in the internal methylation of CoV RNAs. Although the previous study excluded METTL3-mediated RNA (m6A) methylation for lack of the canonical motif RRACH (R and H represent A/G and A/C/T, respectively) (Kim et al., 2020a), we still found many internal methylation sites containing the AAACH motif in the SARS-CoV-2 genome by reanalyzing the Nanopore RNA-seq data. Notable instances include “agTtt” (AAACT on the antisense strand) at the positions 29408 and 29444 (corresponding to the underlined capital letter), and “tgTtt” at the position 29170. Particularly, “tgTtt”, “cgTtt”, “agTtt” and “tgTtt” located at the positions 25402, 26258, 26494 (Figure 4C) and 28235 co-occurred with the NSP15 cleavage motif of four genes (*ORF3a*, *E*, *M* and *N*). In addition, “tgTtt”, “tgTtt”, “ttctT” (the “AAGAA-like” motif on the antisense strand) and “tgTtt” were located at the positions 21566, 21570, 21577 and 21579 (Supplementary Table S1), which are closely linked and flanking the GTTCGT motif of the *S* gene, which merits investigation in the future. The above findings indicated that METTL3 functions in RNA (m6A) methylation of sequences containing the AAACH motif for *ORF3a*, *E*, *M* and *N*, and possibly the “AAGAA-like” motif for *S*, *ORF6*, *7a* and *8*. Finally, we proposed the theoretical arrangement of NSP12-15 and METTL3 in the global RTC structure (Figure 4D) by the integration of information from many aspects, particularly including: 1) identification of NSP15 cleavage sites in our previous study (Li et al., 2021a); 2) discovery of the AAACH motif co-occurred with the NSP15 cleavage motif of four genes; 3) discovery of the association between RNA methylation and

NSP15 cleavage; and 4) discovery of the TRS hairpins of eight genes (*S*, *E*, *M*, *N*, and *ORF3a*, *6*, *7a* and *8*).

By comprehensive analysis of the above results, we constructed a model to answer how the RTC functions in the jumping transcription of CoVs. In this model, the RTC processes double-strand RNAs (dsRNAs) and single-strand RNAs (ssRNAs) in two situations (Figure 4D), respectively. In the first situation, NSP13 unwinds dsRNAs (Yan et al., 2020) to produce ssRNAs(+) or ssRNAs(-), which are processed in two routes. In one route, NSP12 synthesizes RNAs with error correction by NSP14 to produce dsRNAs using unwound ssRNAs(+) or ssRNAs(-) as templates (Knoops et al., 2008). The other route processes ssRNAs(+) or ssRNAs(-), which can be methylated at internal sites and cleaved by NSP15 for jumping transcription. Then, the ssRNAs(+) and ssRNAs(-) are further processed in different ways: most ssRNAs(+) are uncleaved and packaged by the N protein (this is still not clear), while ssRNAs(-) can be uncleaved or cleaved for jumping transcription or degraded, which is regulated by a negative feedback mechanism (Li et al., 2021a). In the second situation, the RTC processes ssRNAs: uncleaved ssRNAs(+) and ssRNAs(-) are used as templates for replication; cleaved ssRNAs(-) are used as templates for transcription. The model can be used to explain the extremely high ratio between sense and antisense reads analyzed in our previous study (Li et al., 2021a) and the experimental result that knockdown of NSP15 by mutation increases the accumulation of viral dsRNA in another previous study (Deng et al., 2017). According to our model, knockdown of NSP15 increases the uncleaved gRNAs(-), which continue to be templates to produce more dsRNAs.

CONCLUSION AND DISCUSSION

In the present study, we proposed the theoretical arrangement of NSP12-15 and METTL3 in the global RTC structure and constructed a model to answer how the RTC functions in the jumping transcription of CoVs. More importantly, our results reveal the complex associations between RNA methylation, NSP15 cleavage, CoV replication and transcription at the molecular level. Our findings enrich fundamental knowledge in the field of gene expression and its regulation, providing a crucial basis for future studies. NSP12-14 form the main structure of the RTC; NSP7 and NSP8, acting as the cofactors of NSP12, may be also included in the main structure of the RTC (Yan et al., 2020); NSP15 and METTL3 are coupled with the main structure. The results of previous experiments suggest that NSP8 is able to interact with NSP15 (Lianqi et al., 2018). Future research needs to be conducted to determine the structures of NSP12&14, NSP12&15, NSP12&METTL3 and NSP15&METTL3 complexes by Cryo-EM. These local RTC structures can be used to assemble a global RTC structure by protein-protein docking calculation. Our model does not rule out the involvement of other proteins (e.g., ORF8) in the global RTC structure or other proteins in the internal methylation of the “AAGAA-like” motif. Future drug design targeting SARS-CoV-2 needs to consider protein-protein and protein-RNA interactions in the RTC, particularly the structure of NSP15 and the TRS hairpin complex.

MATERIALS AND METHODS

The *Betacoronavirus* genus includes five subgenera (*Embecovirus*, *Sarbecovirus*, *Merbecovirus*, *Nobecovirus* and *Hibecovirus*), which were defined as subgroups A, B, C, D and E (Bei et al., 2022). In our previous study (Li et al., 2021b), 1,265 genome sequences of viruses in the *Embecovirus*, *Sarbecovirus*, *Merbecovirus*, *Nobecovirus* subgenera were downloaded from the NCBI Virus database (<https://www.ncbi.nlm.nih.gov/labs/virus>). Two genome sequences (RefSeq: NC_025217 and GenBank: KY352407) of viruses in the *Hibecovirus* subgenus were also downloaded. Among 1,265 genomes, 292 belongs to *Betacoronavirus* subgroup B (including SARS-CoV and SARS-CoV-2). 1,178, 480 and 194 genome sequences of viruses in the *Alphacoronavirus*, *Gammacoronavirus* and *Deltacoronavirus* genera were downloaded to validate the TRS motifs (Figure 1). Nanopore RNA-seq data was downloaded from the website (<https://osf.io/8f6n9/files/>) for reanalysis. Data cleaning and quality control were performed using Fastq_clean (Zhang et al., 2014). Statistics and plotting were conducted using the software R v2.15.3 with the Bioconductor packages (Gao et al., 2014). Protein structure data (PDB: 6 × 1B, 7BQ7 and 7CXN) were used to analyze

REFERENCES

Bei, J., Xu, G., Chang, J., Wang, X., Tung, O. Y., Ruan, J., et al. (2022). SARS-CoV-2 with Transcription Regulatory Sequence Motif Mutation Poses a Greater Threat. *J. South Med. Univ. (In Chin.* 42 (3), 399–404. doi:10.12122/j.jissn.1673-4254.2022.03.12

NSP15, NSP10&16 and NSP7&8&12&13, respectively. The structures of NSP12-16 were predicted using trRosetta (Yang et al., 2020). The minimum free energies (MFEs) of hairpins were estimated by RNAeval v2.4.17 with parameters by manual adjustment.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

AUTHOR CONTRIBUTIONS

SG conceived the project. SG and DM supervised this study. JL, SC, FY and ZC downloaded, managed and processed the data. GD and JS performed programming. XL predicted and analyzed the protein structures. SG drafted the main manuscript text. SG and JR revised the manuscript.

FUNDING

This work was supported by the Yunnan Applied Basic Research—Yunnan Provincial Science and Technology Department—Kunming Medical University joint projects (202101AY070001-073), Shandong Province Natural Science Foundation (ZR2020QC071) to GD and National Natural Science Foundation of China (31900444) to ZC.

ACKNOWLEDGMENTS

We are grateful for the help from the following faculty members of College of Life Sciences at Nankai University: Qiang Zhao, Jia Chang, Jianyi Yang and Bingjun He and that from Jinlong Bei and Tung On Yau. We would like to thank Editage (www.editage.cn) for polishing part of this manuscript in English language. This manuscript was online as a preprint on 18 Feb 2021 at <https://biorxiv.org/cgi/content/short/2021.02.17.431652v1>.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2022.904513/full#supplementary-material>

Deng, X., Hackbart, M., Mettelman, R. C., O'Brien, A., Mielech, A. M., Yi, G., et al. (2017). Coronavirus Nonstructural Protein 15 Mediates Evasion of dsRNA Sensors and Limits Apoptosis in Macrophages. *Proc. Natl. Acad. Sci. U. S. A.* 114 (21), E4251–E4260. doi:10.1073/pnas.1618310114

Denison, M. R., Graham, R. L., Donaldson, E. F., Eckerle, L. D., and Baric, R. S. (2011). Coronaviruses. *RNA Biol.* 8 (2), 270–279. doi:10.4161/rna.8.2.15013

- Duan, G., Shi, J., Xuan, Y., Chen, J., Liu, C., Ruan, J., et al. (2020). 5' UTR Barcode of the 2019 Novel Coronavirus Leads to Insights into its Virulence. *Chin. J. Virology (In Chinese)* 36 (3), 365–369. doi:10.13242/j.cnki.bingduxuebao.003681
- Gao, S., Ou, J., and Xiao, K. (2014). *R Language and Bioconductor in Bioinformatics Applications*. Chinese Edition. Tianjin: Tianjin Science and Technology Translation Publishing Ltd.
- Grossoehme, N. E., Li, L., Keane, S. C., Liu, P., Dann, C. E., Leibowitz, J. L., et al. (2009). Coronavirus N Protein N-Terminal Domain (NTD) Specifically Binds the Transcriptional Regulatory Sequence (TRS) and Melts TRS-cTRS RNA Duplexes. *Journal of molecular biology* 394, 544–557. doi:10.1016/j.jmb.2009.09.040
- Hillen, H. S., Kokic, G., Farnung, L., Dienemann, C., Tegunov, D., and Cramer, P. (2020). Structure of Replicating SARS-CoV-2 Polymerase. *Nature* 584 (7819), 154–156. doi:10.1038/s41586-020-2368-8
- Jiayuan, C., Jinsong, S., Yau Tung, O., Chang, L., Xin, L., Qiang, Z., et al. (2020). Bioinformatics Analysis of the 2019 Novel Coronavirus Genome. *Chinese Journal of Bioinformatics (In Chinese)* 18 (2), 96–102. doi:10.12113/202001007
- Kim, D., Lee, J.-Y., Yang, J.-S., Kim, J. W., Kim, V. N., and Chang, H. (2020). The Architecture of SARS-CoV-2 Transcriptome. *Cell* 181 (4), 914–921. doi:10.1016/j.cell.2020.04.011
- Kim, Y., Jedrzejczak, R., Maltseva, N. I., Wilamowski, M., Endres, M., Godzik, A., et al. (2020). Crystal Structure of Nsp15 Endoribonuclease NendoU from SARS-CoV-2. *Protein Science* 29 (7), 1596–1605. doi:10.1002/pro.3873
- Knoops, K., Kikkert, M., Worm, S. H. E. v. d., Zevenhoven-Dobbe, J. C., van der Meer, Y., Koster, A. J., et al. (2008). SARS-coronavirus Replication Is Supported by a Reticulovesicular Network of Modified Endoplasmic Reticulum. *PLoS Biol* 6 (9), e226. doi:10.1371/journal.pbio.0060226
- Krafcikova, P., Silhan, J., Nencka, R., and Boura, E. (2020). Structural Analysis of the SARS-CoV-2 Methyltransferase Complex Involved in RNA Cap Creation Bound to Sinefungin. *Nat Commun* 11 (1), 3717. doi:10.1038/s41467-020-17495-9
- Li, X., Duan, G., Zhang, W., Shi, J., Chen, J., Chen, S., et al. (2020). A Furin Cleavage Site Was Discovered in the S Protein of the 2019 Novel Coronavirus. *Chinese Journal of Bioinformatics (In Chinese)* 18 (2), 103–108. doi:10.12113/202002001
- Li, X., Chang, J., Chen, S., Wang, L., Tung, O. Y., Zhao, Q., et al. (2021). Genomic Feature Analysis of Betacoronavirus Provides Insights into SARS and COVID-19 Pandemics. *Frontiers Microbiology* 10, 1–11. doi:10.3389/fmicb.2021.614494
- Li, X., Cheng, Z., Wang, F., Chang, Jia., Zhao, Q., Zhou, H., et al. (2021). A Negative Feedback Model to Explain Regulation of SARS-CoV-2 Replication and Transcription. *Frontiers in Genetics* 10, 1–11. doi:10.3389/fgene.2021.641445
- Lianqi, Z., Lei, L., Liming, Y., Zhenhua, M., Zhihui, J., Zhiyong, L., et al. (2018). Structural and Biochemical Characterization of Endoribonuclease Nsp15 Encoded by Middle East Respiratory Syndrome Coronavirus. *Journal of Virology* 92(22):e00893-18. doi:10.1128/JVI.00893-18
- Liu, C., Chen, Z., Hu, Y., Ji, H., Yu, D., Shen, W., et al. (2018). Complemented Palindromic Small RNAs First Discovered from SARS Coronavirus. *Genes (Basel)* 9 (9), 1–11. doi:10.3390/genes9090442
- Mihindukulasuriya, K. A., Wu, G., St. Leger, J., Nordhausen, R. W., and Wang, D. (2008). Identification of a Novel Coronavirus from a Beluga Whale by Using a Panviral Microarray. *J Virol* 82, 5084–5088. doi:10.1128/jvi.02722-07
- Papineau, A., Berhane, Y., Wylie, T. N., Wylie, K. M., Sharpe, S., and Lung, O. (2019). Genome Organization of Canada Goose Coronavirus, A Novel Species Identified in a Mass Die-Off of Canada Geese. *Sci Rep* 9, 5954. doi:10.1038/s41598-019-42355-y
- Sawicki, S. G., and Sawicki, D. L. (1998). “A New Model for Coronavirus Transcription,” in *Coronaviruses and Arteriviruses*. Editors L. Enjuanes, S. G. Siddell, and W. Spaan (Boston, MA: Springer US), 215–219. doi:10.1007/978-1-4615-5331-1_26
- Silva, S. J. R., Alves da Silva, C. T., Mendes, R. P. G., and Pena, L. (2020). Role of Nonstructural Proteins in the Pathogenesis of SARS-CoV-2. *J Med Virol* 92, 1427–1429. doi:10.1002/jmv.25858
- Sola, I., Moreno, J. L., Zúñiga, S., Alonso, S., and Enjuanes, L. (2005). Role of Nucleotides Immediately Flanking the Transcription-Regulating Sequence Core in Coronavirus Subgenomic mRNA Synthesis. *J Virol* 79, 2506–2516. doi:10.1128/JVI.79.4.2506-2516.2005
- Xu, X., Ji, H., Jin, X., Cheng, Z., Yao, X., Liu, Y., et al. (2019). Using pan RNA-Seq Analysis to Reveal the Ubiquitous Existence of 5' and 3' End Small RNAs. *Front Genet* 10, 105–111. doi:10.3389/fgene.2019.00105
- Yan, L., Zhang, Y., Ge, J., Zheng, L., Gao, Y., Wang, T., et al. (2020). Architecture of a SARS-CoV-2 Mini Replication and Transcription Complex. *Nat Commun* 11 (2020), 5874–5876. doi:10.1038/s41467-020-19770-1
- Yang, J., Anishchenko, I., Park, H., Peng, Z., Ovchinnikov, S., and Baker, D. (2020). Improved Protein Structure Prediction Using Predicted Interresidue Orientations. *Proc. Natl. Acad. Sci. U.S.A.* 117 (3), 1496–1503. doi:10.1073/pnas.1914677117
- Yount, B., Roberts, R. S., Lindesmith, L., and Baric, R. S. (2006). Rewiring the Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) Transcription Circuit: Engineering a Recombination-Resistant Genome. *Proc. Natl. Acad. Sci. U.S.A.* 103 (33), 12546–12551. doi:10.1073/pnas.0605438103
- Zhang, M., Zhan, F., Sun, H., Gong, X., Fei, Z., and Gao, S. (2014). “Fastq_clean: An Optimized Pipeline to Clean the Illumina Sequencing Data with Quality Control,” in Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Belfast, UK, 2-5 Nov. 2014, 44–48. doi:10.1109/BIBM.2014.6999309

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Liang, Shi, Chen, Duan, Yang, Cheng, Li, Ruan, Mi and Gao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.