Genes & Genomics (2022) 44:1399–1404
https://doi.org/10.1007/s13258-022-01299-w

Online ISSN 2092-9293
Print ISSN 1976-9571

RESEARCH ARTICLE

Discovery of a tRNA-like base sequence in the coronavirus genome and possible mechanism of action

Hachiro Inokuchi[1]

Received: 28 April 2022 / Accepted: 25 July 2022 / Published online: 27 September 2022

Abstract

Background The question of whether the coronavirus genome contain as-yetununderstood genetic component.
Purpose (Objective) Elucidate the novel functions of the discovered tRNA-like base sequence and lead to the development of novel therapeutic agents.
Methods A novel tRNA-like base sequence was found in the sequences complementary to the genomes of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and SARS-CoV. By comparing mutations in the tRNA-like base sequences of these two viruses, it was found that base pairing in the cloverleaf model of SARS-CoV-2 was more robust than that of SARS-CoV.
Results The results of homology search between a short sequence of the coronavirus tRNA-like base sequence and human genes suggest that the molecule produced by this novel tRNA-like sequence may be involved in the splicing of human messenger RNA.
Conclusions Experimental molecular evidence of the tRNA-like base sequence discovered in this study is urgently needed.

Keywords Coronavirus · Cloverleaf model of tRNA · Mutation · Base sequence · Intron

Introduction

Infection with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is now widespread, and the associated coronavirus disease (COVID-19) has become a pandemic worldwide. By December 3, 2021, 199,936,878 cases were confirmed in more than 110 countries, with 4,254,571 deaths (Elflein 2021). SARS-CoV-2 is the seventh coronavirus known to infect humans (Coronaviridae Study Group of the International Committee on Taxonomy of Viruses 2020; Wu et al. 2020; Zhou et al. 2020). Among these, SARS-CoV, Middle East respiratory syndrome coronavirus (MERS-CoV), and SARS-CoV-2 can cause severe disease, whereas human coronaviruses HKU1, NL63, OC43, and 229E are associated with mild symptoms (Coronaviridae Study Group of the International Committee on Taxonomy of Viruses 2020; V'kovski et al. 2021). In this perspective study, comparative analysis of genomic data was performed to deduce the characteristics of the SARS-CoV-2 genome. This revealed the notable features of a novel SARS-CoV-2 tRNA sequence, whose mechanism of action was also hypothesized.

Materials and methods

A 330-base tRNA-like sequence was visually identified in the complementary strand of the SARS-CoV-2 genome without using any software such as tRNA Scan-SE. This method was based on the detection of the T-loop sequence (G/CNNNRANNC/G) and on subsequent examination of the following stem structure (T-stem). Next, the sequence was extended in the 3′ direction to search for amino acid acceptor stems (AA-stem). The analysis tool "Analyze this sequence"/"Find in this sequence" included in the NCBI gene database was used to detect both of these sequences. Meanwhile, an ochre anticodon sequence was searched at a position approximately 20 bases away from the T-stem in the 5′ direction, and it was investigated whether an anticodon stem of five bases could be assembled. Finally, a D-loop/stem and an intron were placed, several cloverleaf structure

✉ Hachiro Inokuchi
h-ino@maia.eonet.ne.jp

[1] Yoshida Biological Laboratory Inc., 11-1, Takehanasotoda-cho, Yamashina-ku, Kyoto 607-8081, Japan

models were drawn one by one, and the best structural model was selected.

Homology search between short sequences (eight bases) of coronavirus and human genes (one million bases or less) was conducted using the same "Analyze this sequence"/"Find in this sequence" tool, and the number and position of the highlighted sequences that appeared on the PC screen were annotated. For all analyses, a Toshiba Dynabook Satellite L5-C Series PC was used.

Results

The cloverleaf model of the novel tRNA-like base sequence, consisting of 330 bases, as well as its position (12,940–13,296) on the genome of SARS-CoV-2 (NCBI/Nucleotide/NC_045512.2 complement) are shown in Fig. 1. Although the cloverleaf model, showing a stem and loop structure, is consistent with the conventional model (Dirheimer et al. 1995), it has the following peculiar characteristics: (1) it contains an intron consisting of 260 bases between the D-stem and the AC-stem; (2) its anticodon is TTA, corresponding to the ochre stop codon; and (3) although the number of bases in its T-loop, consisting of seven bases, matches that of conventional T-loops, the static base sequence $G_{53}TTC_{56}$ is substituted with $C_{53}TTT_{56}$.

The mutations present in the genome of the new coronavirus SARS-CoV-2 at bases 16,635–16,964, based on 4787 sequenced deposited in the NCBI database as of March 25, 2021, are shown in Fig. 2a. After identification of these mutations, this region of the SARS-CoV-2 genome was compared with that of the previously widespread Tor2 isolate of SARS-CoV (NCBI/Nucleotide/NC_004718.3) (Fig. 2b). Fifty-two mutations were identified in the 330-base tRNA-like sequence. The comparison of exons and introns of the 330-base tRNA-like sequences of different SARS-CoV-2 isolates revealed a smaller number of mutations in the exon region. Presumably, this depends on the bias of exons forming the secondary structure of the tRNA in the cloverleaf model, suggesting the validity of the assumed stem structure of the discovered tRNA base sequence. Conversely, the comparison of the tRNA structures of SARS-CoV and SARS-CoV-2 strains pointed at five mutations in the stem region. These mutations are in the direction of stabilizing base pairing in the stem, and there are no reverse mutations. In addition, a mutation (C→T) at position 33 of the constant base is consistent with the general model of the tRNA structure. These results indicate that the increased pathogenicity

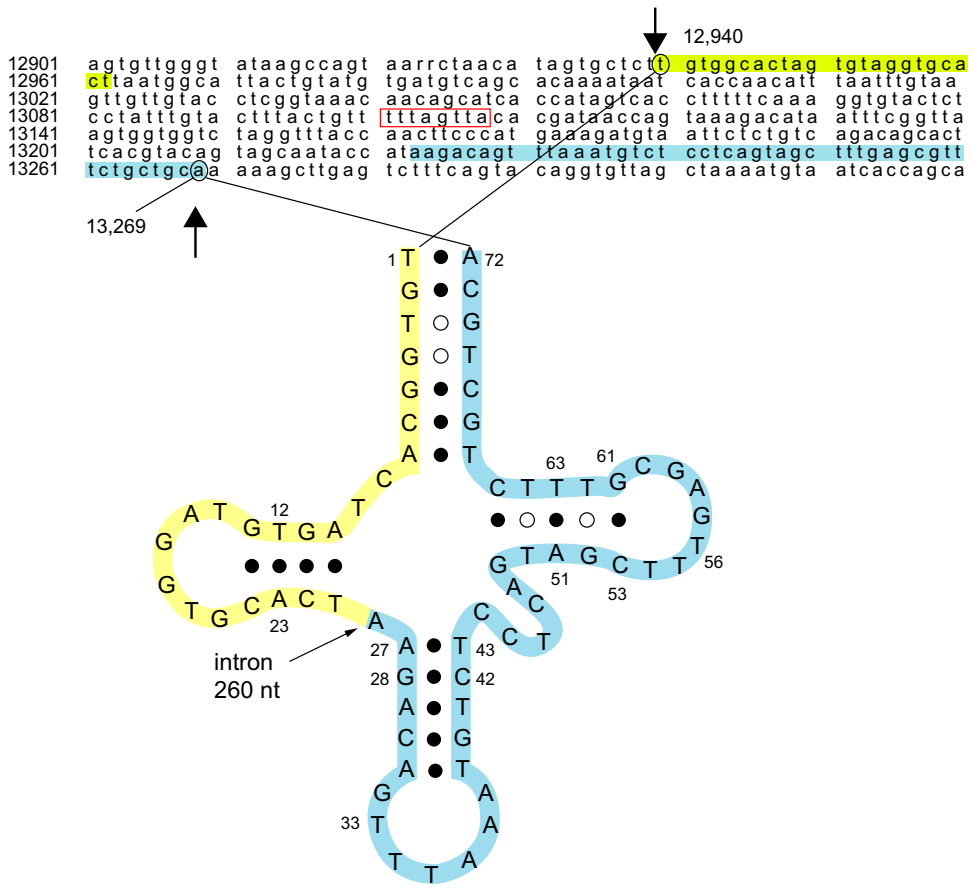

Fig. 1 Cloverleaf model of the 330-base tRNA-like sequence found in the genome of the new coronavirus SARS-CoV-2. The tRNA-like base sequence occupies 330 bases at positions 12,940–133,269 of the NCBI/Nucleotide/NC_045512-2 genome (complement). It contains an intron consisting of 260 bases. The exon regions are highlighted in color (orange, 5′ side and blue, 3′ side). The eight bases in the sequence chart (TTTAGTTA), surrounded by a red line, represent the central part of the 330-base tRNA-like sequence, which was used for homology search. Although U should be used in the cloverleaf model of tRNA, T was used for the purpose of sequence description (● indicates Watson-Crick-type pairing, and ◯ indicates non-Watson-Crick base pairing)

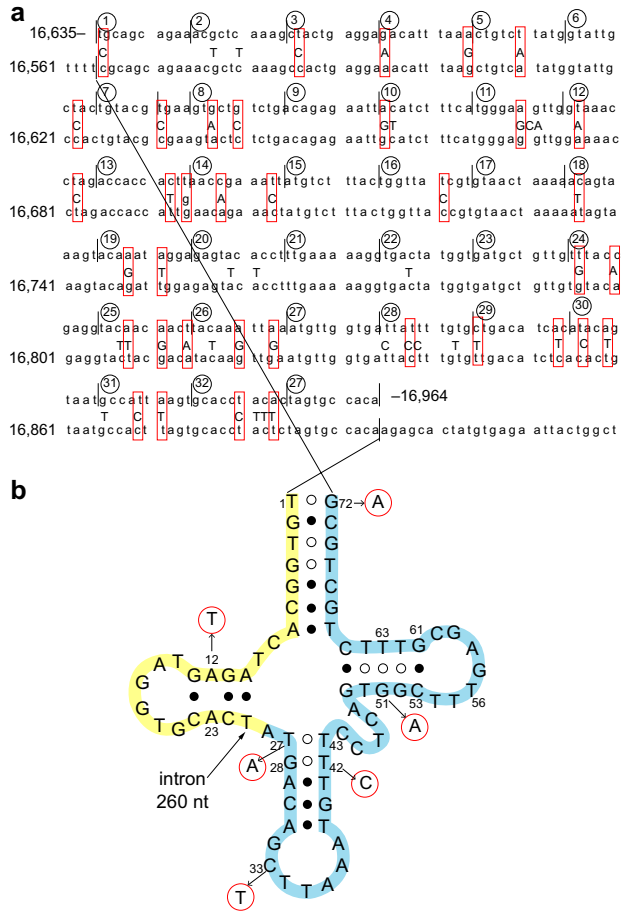

Fig. 2 Mutations found in the 330-base tRNA-like sequence. **a** The base sequence starting at position 16,635 represents the 330-base tRNA-like sequence of the new coronavirus [NCBI/Nucleotide/NC_045512.2, (+) strand], while the base sequence starting at position 16,561 represents the 330-base tRNA-like sequence of SARS-CoV Tor2 [NCBI/Nucleotide /NC_004718.3, (+) strand]. Mutated bases are shown in capital letters between the two base sequences. Bases differing between SARS-CoV-2 (upper base sequence) and SARS-CoV (lower base sequence) are surrounded by a red line. ①–㉝ indicate the section number of the 330-base tRNA-like sequence. **b** Cloverleaf model of the 330-base tRNA-like sequence of SARS-CoV Tor2. Note that the model is drawn from complementary strands. →◯ indicates a mutation in the novel strain, SARS-CoV-2 (Fig. 1)

in the new coronavirus strain compared to the previous strain is partially proportional to the stability of its tRNA structure.

To obtain information about the possible validity and importance of the novel tRNA-like base sequence, a homology search between this 330-base tRNA-like sequence and human genes was carried out. The 330-base tRNA-like sequence was divided into 33 sections [◯ in Fig. 2a], and four fragments with different complementarity and orientation were generated in each section. The results of screening of 10 large and small olfactory receptor genes using these eight-base fragments are shown in Supplementary Table S1. Considering these results, the fragment no. 4 of section ⑰ (ATTGATTT) was selected for further investigation. This choice was made based on the large number of homologous sequences to this fragment, which filled up a whole PC monitor; the absence of mutations in this section; and the fact that section ⑰ is located exactly in the center of the 330-base tRNA-like sequence.

The position of an A base at the 5′ end of fragment no. 4 of section ⑰ (ATTGATTT) on the human genes, as well as the position of the codon corresponding to the TTA anticodon located upstream, are shown in Fig. 3a. The number of homologous hits is almost proportional to the size of the gene, and homologous sequences are scattered within the gene. When focusing on the interval between these sequences, it was found that the ATTGATTT sequence exists in pairs at a constant distance (461 ± 2 bases) in some human olfactory receptor genes, and that there is always a codon corresponding to the TTA anticodon located upstream at a fixed position of 120 ± 3 bases. The data for seven genes, including *SFTPD*, *PGL*, and *IL7*, out of approximately 150 genes obtained by searching "*Homo sapiens* genes, COVID-19" in the NCBI database, are shown in Fig. 3.

Although the 5′ATTGATTT3′ sequence is often found in human genes, only the 3′ATTGATTT5′ sequence is found in section ⑰ of the coronavirus genome, and these two sequences do not pair by Watson–Crick complementarity even if they are arranged side by side. Interestingly, however, these sequences can form complementary pairs by shifting one of the sequences by two bases to the left or right and considering the non-Watson–Crick G–T and T–T base pairs, which are often found in the cloverleaf model of tRNA (Clark 1979; Ozeki et al. 1987).

The position of the ATTGATTT sequence, existing in pairs, and that of the stop codons located upstream thereof on the base sequence of the human olfactory receptor gene *OR2L13* are shown in Fig. 3b. As an example, it is shown that these sequences can undergo complementary pairing by shifting two bases to the right (Fig. 4). To carry out pairing by two-base shift, one of the two bases adjacent to the ends of ATTGATTT in human genes must not be C. Notably, all of the ATTGATTT—(461 ± 2 bases)—ATTGATTT sequences illustrated in Fig. 3 do not have C in at least one of the two positions adjacent to the ends of ATTGATTT.

Despite its relevance within the scope of the study, the ATTGATTT—(461 ± 2 bases)—ATTGATTT sequences found in human genes were not located in the exons, but rather in the introns. On the other hand, considering all eight-base fragments of the 330-base tRNA-like sequence, no other eight-base sequence was found to appear in pairs at regular intervals except for the ATTGATTT sequence from the central part of the tRNA-like base sequence.

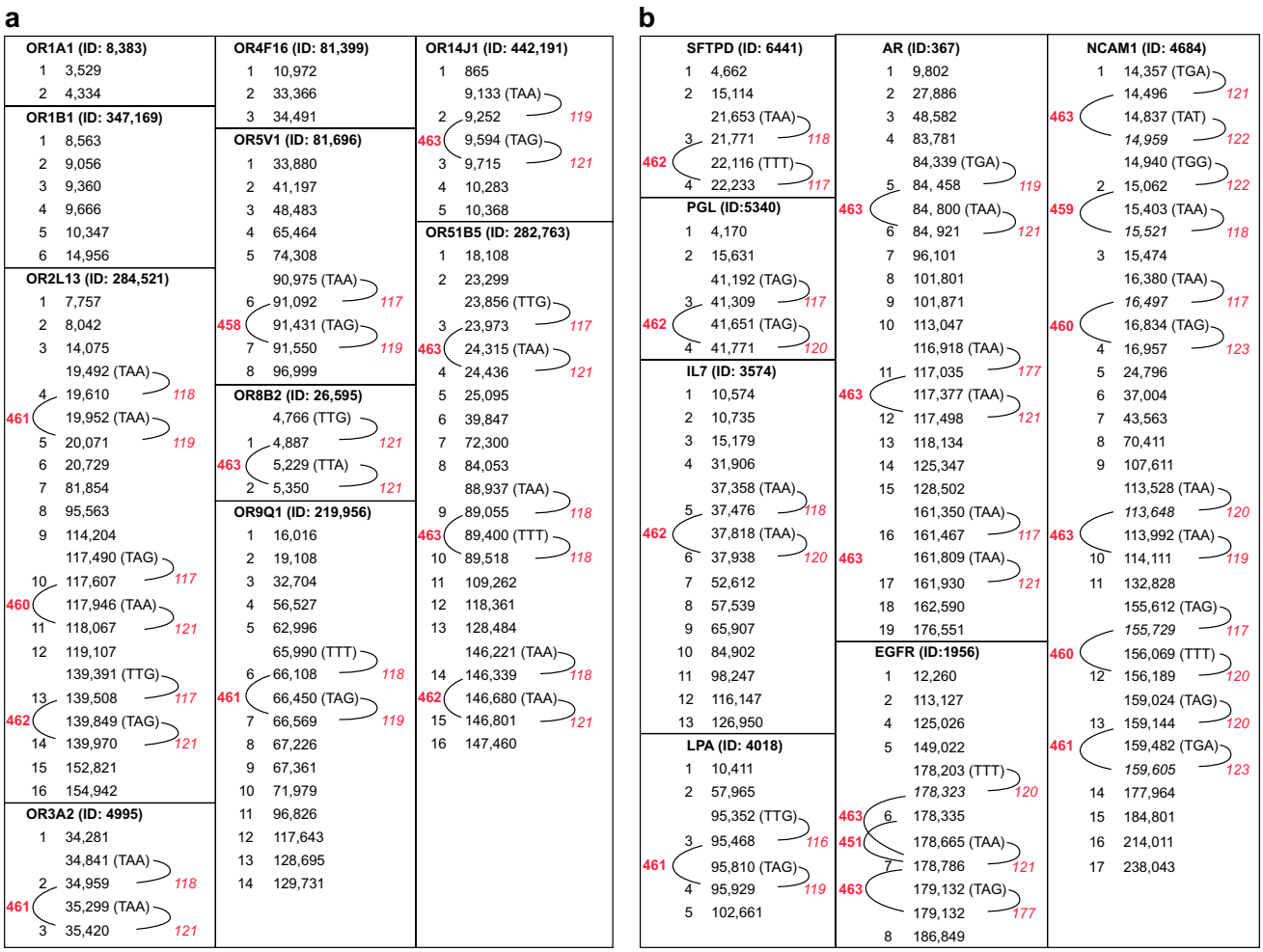

Fig. 3 Positions and intervals of the eight-base fragment ATTGATTT and the stop codon. **a** Human olfactory receptor genes; **b** COVID-19-related human genes. Numbers in each gene name indicate their order of discovery. The numerical values on the right side indicate the position of each gene in the genome. The numbers in bold red on the semi-elliptical line indicate the distance (number of bases) between the eight-base fragments. Stop codons are enclosed in parentheses at their respective positions. The distance (number of bases) between the eight-base fragment and the stop codon is indicated by red numbers in italics

Discussion

The above results allowed to hypothesize the mechanism of action of the 330-base tRNA-like sequence on human genes (Fig. 5). Splicing of messenger RNA is essential for the normal function of human genes containing introns (Sharp 1994); this reaction involves spliceosomes, which bind to the intron regions (Warf and Berglund 2010; Will and Lührmann 2011). However, if the intron of a human gene contains a pair of ATTGATTT sequences spaced 461 ± 2 bases apart and an upstream codon corresponding to the ochre anticodon at a distance of 120 ± 3 bases, two 330-base tRNA-like molecules may form a dimer by complementary pairing with the intron of messenger RNA. This process would interfere with the binding of spliceosomes and thus impair splicing of messenger RNA. Such interference would cause fatal damage to gene expression, leading to a deficiency in gene function.

According to the model, the tRNA structural base sequence (intron sequence) contains functional sites similar to those of transfer-messenger RNA (tmRNA) found in prokaryotes (Keiler 2008; Komine et al. 1994; Ushida et al. 1994); in particular, the intron does not need to be

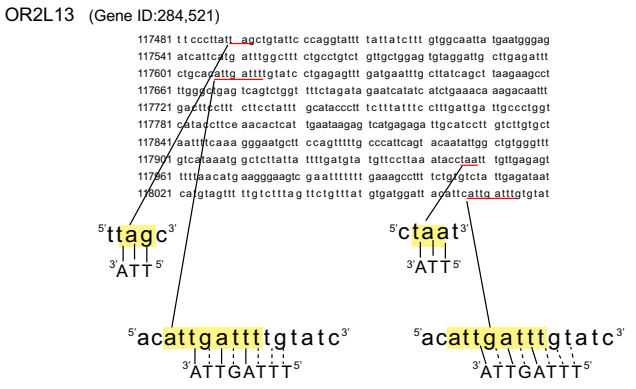

Fig. 4 Position of a pair of eight-base sequences with regular intervals in the base sequence of the human olfactory receptor gene *OR2L13* (NCBI/Gene ID: 284521), and positions of stop codons located upstream of the pair of eight-base sequences. These sequences are highlighted in the gene sequence chart, and complementary pairing is shown with the TTA anticodon and the eight-base sequence from the 330-base tRNA-like sequence of SARS-CoV-2 (using uppercase letters). Complementary bonds of the Watson-Crick type, solid lines; non-Watson-Crick type, dotted lines

spliced out of the tRNA base sequence for the latter to be functional.

The genes described in this study were identified manually through an informatics search, focusing only on a few genes thought to be related to SARS-CoV-2 infection. More genes will likely be identified by conducting a similar homology analysis on genes and short segment/fragment base sequences in the human genome using large-scale and modern information processing technology. Although it is unlikely that all these genes are involved in coronavirus infection, it will be surely possible to identify groups of genes that cause exacerbation and sequelae of infection by combining different research approaches. In any case, experimental molecular evidence of the tRNA-like base sequence discovered in this study is urgently needed. I firmly believe that basic molecular biology research on tRNA genes will be one of the driving forces behind humankind's victory in the fight against the coronavirus.

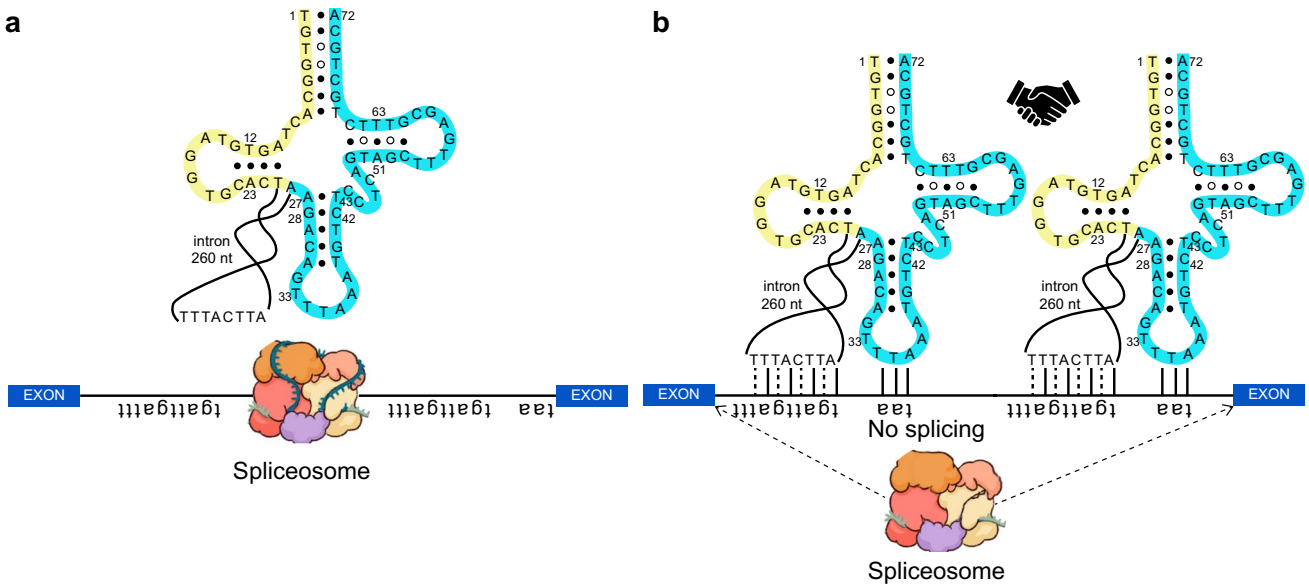

Fig. 5 Explanatory diagram of the hypothetical mechanism of action of the 330-base tRNA-like sequence on the splicing of messenger RNA of human genes. **a** The spliceosome binds to the intron and splicing occurs normally; however, **b** when a dimeric molecule of the 330-base tRNA-like sequence binds to the intron by complementary binding, the spliceosome cannot bind to the intron, and splicing does not occur

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s13258-022-01299-w.

Acknowledgements The author acknowledges Dr. Akira Wada and Dr. Chieko Wada, from the Yoshida Biological Laboratory, for helpful discussions and their encouragement. The author also recognizes the editorial support, in the form of medical writing, table assembly, creation of high-resolution images based on the author's detailed instructions, collation of author comments, copyediting, fact checking, and referencing, that was provided by Editage, Cactus Communications.

Author contributions Not applicable.

Funding Not applicable.

Availability of data and materials All data generated or analyzed during this study are included in this published article and its supplementary information file.

Declarations

Conflict of interest The author declares no competing interests.

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

Research involving human participants and/or animals This study did not involve human participants or animals.

Informed consent Not applicable.

References

Clark BFC (1979) Structure and function of tRNA. In: Celis JE, Smith JD (eds) Nonsense mutations and tRNA suppressors. Academic Press, London, pp 1–46

Coronaviridae Study Group of the International Committee on Taxonomy of Viruses (2020) CSG of the international. The species severe acute respiratory syndrome-related coronavirusnCOV and naming it SARS-COV-2a. Nat Microbiol 5:536–544

Dirheimer G, Keith G, Dumas P, Westhof E (1995) Primary, secondary, and tertiary structure of tRNAs. In: Soll D, Rajbhandary U (eds) tRNA: structure, biosynthesis, and functions. American Society for Microbiology, Washington, DC, pp 93–126

Elflein J (2021) Coronavirus (COVID 19) disease pandemic—statistics and facts. Statista. https://www.statista.com/topics/5994/the-coronavirus-disease-covid-19-outbreak/#dossierKeyfigures. Accessed 25 Feb 2022.

Keiler KC (2008) Biology of trans-translation. Annu Rev Microbiol 62:133–151

Komine Y, Kitabatake M, Yokogawa T, Nishikawa K, Inokuchi H (1994) A tRNA-like structure is present in 10Sa RNA, a small stable RNA from *Escherichia coli*. Proc Natl Acad Sci USA 91:9223–9227

Ozeki H, Ohyama K, Inokuchi H et al (1987) Genetic system of chloroplasts. Cold Spring Harb Symp Quant Biol 52:791–804

Sharp PA (1994) Split genes and RNA splicing. Cell 77:805–815

Ushida C, Himeno H, Watanabe T, Muto A (1994) tRNA-like structures in 10Sa RNAs of *Mycoplasma capricolum* and *Bacillus subtilis*. Nucleic Acids Res 22:3392–3396

Vkovski P, Kratzel A, Steiner S, Stalder H, Thiel V (2021) Coronavirus biology and replication: implications for SARS-CoV-2. Nat Rev Microbiol 19:155–170

Warf MB, Berglund JA (2010) Role of RNA structure in regulating pre-mRNA splicing. Trends Biochem Sci 35:169–178

Will CL, Lührmann R (2011) Spliceosome structure and function. Cold Spring Harb Perspect Biol 3:a003707

Wu F, Zhao S, Yu B et al (2020) A new coronavirus associated with human respiratory disease in China [published correction appears in Nature 2020;580(7803):E7]. Nature 579:265–269

Zhou P, Yang XL, Wang XG et al (2020) A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 579:270–273