genes

MDPI

Article

Genomic Analysis of Non-B Nucleic Acids Structures in SARS-CoV-2: Potential Key Roles for These Structures in Mutability, Translation, and Replication?

Stefan Bidula [1,*] and Václav Brázda [2]

[1] School of Pharmacy, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, UK
[2] Institute of Biophysics of the Czech Academy of Sciences, 61265 Brno, Czech Republic
* Correspondence: s.bidula@uea.ac.uk

Abstract: Non-B nucleic acids structures have arisen as key contributors to genetic variation in SARS-CoV-2. Herein, we investigated the presence of defining spike protein mutations falling within inverted repeats (IRs) for 18 SARS-CoV-2 variants, discussed the potential roles of G-quadruplexes (G4s) in SARS-CoV-2 biology, and identified potential pseudoknots within the SARS-CoV-2 genome. Surprisingly, there was a large variation in the number of defining spike protein mutations arising within IRs between variants and these were more likely to occur in the stem region of the predicted hairpin stem-loop secondary structure. Notably, mutations implicated in ACE2 binding and propagation (e.g., ΔH69/V70, N501Y, and D614G) were likely to occur within IRs, whilst mutations involved in antibody neutralization and reduced vaccine efficacy (e.g., T19R, ΔE156, ΔF157, R158G, and G446S) were rarely found within IRs. We also predicted that RNA pseudoknots could predominantly be found within, or next to, 29 mutations found in the SARS-CoV-2 spike protein. Finally, the Omicron variants BA.2, BA.4, BA.5, BA.2.12.1, and BA.2.75 appear to have lost two of the predicted G4-forming sequences found in other variants. These were found in *nsp2* and the sequence complementary to the conserved stem-loop II-like motif (S2M) in the 3′ untranslated region (UTR). Taken together, non-B nucleic acids structures likely play an integral role in SARS-CoV-2 evolution and genetic diversity.

Keywords: SARS-CoV-2; inverted repeats; G-quadruplex; pseudoknot; spike protein; mutation; adaptation

check for updates

Citation: Bidula, S.; Brázda, V. Genomic Analysis of Non-B Nucleic Acids Structures in SARS-CoV-2: Potential Key Roles for These Structures in Mutability, Translation, and Replication? *Genes* **2023**, *14*, 157. https://doi.org/10.3390/genes14010157

Academic Editors: Emanuela Ruggiero, Marco Di Antonio and Samantha Kendrick

Received: 9 December 2022
Revised: 1 January 2023
Accepted: 4 January 2023
Published: 6 January 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

When we consider the structure of nucleic acids, our first thoughts are of the iconic DNA beta-helical structure. However, nucleic acids are structurally diverse and can be found in a wide range of topologies and conformations within both living and non-living entities. Non-B nucleic acids have been identified as important regulators in fundamental biological processes and have emerged as novel therapeutic targets within infection and disease.

There is growing evidence that these non-canonical nucleic acids structures, such as G-quadruplexes (G4s), cruciforms, hairpins, and pseudoknots, may contribute to both the functional biology and mutational variability of humans, animals, plants, and microorganisms [1–5]. Inverted repeats (IRs) constitute a sequence of nucleotides followed downstream by its reverse completed sequence, often separated by a 'loop' sequence. Viral origins of replication and bacterial plasmids are found to be enriched with IRs [6,7]. These IR sequences can fold into a hairpin stem-loop structure or palindrome in single-stranded nucleic acids. This can significantly contribute to genomic instability and mutation [8]. Furthermore, they have been implicated in a wide range of biological processes, such as replication, transcription, and DNA repair [9,10]. IRs also regulate RNA processing in animals and plants, and transcripts containing IRs are processed to produce small RNAs which silence genes [11,12]. IRs are also an important component of pseudoknots: a common structural motif in RNA formed of two nested stem-loops [13]. Pseudoknots have

been found to be present in viruses whereby they contribute to viral translation, replication, and can also induce frameshifts [5]. Pseudoknots have also been shown to act as binding sites for proteins and may act as regulatory switches in response to environmental signals. They are highly conserved amongst viruses, and, as such, they are beginning to emerge as a potential antiviral target for SARS-CoV-2 [14].

G4s are four-stranded nucleic acids structures that arise in guanine-rich regions of RNA/DNA, and are formed in sequences composed of four runs of ≥two guanines separated by a nucleotide loop (e.g., **GG**AT**GG**AT**GG**AT**GG**) [15]. Here, four guanines associate via Hoogsteen hydrogen bonding to form a G-tetrad. These G-tetrads stack upon one another and are stabilised by a metal cation (e.g., K^+) to form the G4 secondary structure. These structures have been gaining interest recently as antimicrobial targets, due to their demonstrable roles in the regulation of fundamental biological processes such as transcription, translation, replication, and alternative gene splicing [15]. Indeed, G4s have arisen as promising drug targets within bacteria, viruses, parasites, and fungi [16–19].

Goswami and colleagues recently highlighted that SARS-CoV-2 hot-spot mutations were significantly enriched within IRs in the Wuhan reference genome, and hypothesised that IRs could contribute to further mutational drive [20]. This hypothesis was confirmed in additional variants, but in-depth analyses of IRs in more recently identified variants have not been conducted [21]. Moreover, G4s have recently arisen as promising targets to treat SARS-CoV-2 infections [22]. The important roles of these non-canonical nucleic acids structures in SARS-CoV-2 are only just starting to become apparent. Thus, critical biological insights into the roles these structures may have in SARS-CoV-2 could help with understanding the biology of this virus and unveil novel druggable targets to treat these infections. In this article, we analyse SARS-CoV-2 genomes for the presence of IRs, pseudoknots, and G4s, with the aim of stimulating new schools of thought and identifying future experimental directions for the fields of nucleic acids biology and virology.

2. Materials and Methods

2.1. Selection of Sequences

Representative genomes for the currently circulating variant of concern (Omicron), formerly circulating variants of concern (Alpha, Beta, Gamma, and Delta), 9 formerly monitored variants (Epsilon, 20A, Kappa, Iota, 20B, Eta, Theta, Lambda, and Mu), and the Wuhan reference strain were analysed. The FASTA sequences for the entire genomes and the *S* genes encoding the SARS-CoV-2 spike glycoproteins for each were obtained from the National Center for Biotechnology Information (NCBI; last accessed 19 December 2022). Only complete genomes and sequences were used for analysis. Representative genomes were used as there was negligible variation between the locations of predicted non-B structures amongst all genomes from the same variant. The accession numbers of the genomes analysed can be found in Table S1 and the genome information can be found in the Supplementary Materials.

2.2. Detection of Mutations within IRs, Prediction of Pseudoknot Formation, and G4-Analysis

To quantify the number of predicted IRs within the *S* genes, the FASTA sequences were analysed using the Palindrome Analyser web server (http://palindromes.ibp.cz/#/en/index; (last accessed on 19 December 2022) [23]) using the default settings (size: 6–30 bp, spacer: 0–10 bp, and mismatches: 0, 1). Defining, shared, and unique mutations were identified via CoVariants (https://covariants.org/; (last accessed on 19 December 2022) [24]), which collates raw data provided by the Global Initiative on Sharing All Influenza Data ((GISAID); [25]). Prior to post-analysis, FASTA sequences of the variants' *S* genes were aligned to the Wuhan reference sequence using Clustal Omega (EMBL-EBI) to account for any effects of the deletion mutations and differences in nucleotide number. Mutations were noted to have occurred within an IR only if the mutation site fell within the stem or loop region of the predicted IR. Pseudoknot formation was predicted using ProbKnot within the RNAstructure program as described previously [26,27]. Pseudoknot predictions were

performed using 1 iteration and a minimum helix length of 3. The ProbKnot CT files containing the predicted pseudoknot structures are provided in the Supplementary Materials. The presence of G4-forming sequences in the SARS-CoV-2 genomes was determined via QGRS Mapper using the search options of max length = 30, minimum group size = 2, and loop size = 0–12 [28].

2.3. Statistical Analysis

Data comparing groups were first tested for normality via a Shapiro–Wilk normality test prior to analysis via either an unpaired Student's *t*-test or one-way ANOVA depending upon the number of variables. Significance was given as any value < 0.05.

3. Results

3.1. There Is a Large Variation in the Number of Defining Mutations Falling within IRs between SARS-CoV-2 Variants

In-depth analyses of IRs in more recently identified SARS-CoV-2 variants have not been conducted [21]. Therefore, we first identified the presence of IRs in the entire genome and *S* genes of the currently circulating variant of concern (Omicron), formerly circulating variants of concern (Alpha, Beta, Gamma, and Delta), and nine formerly monitored variants (Epsilon, 20A, Kappa, Iota, 20B, Eta, Theta, Lambda, and Mu) to offer insight into whether SARS-CoV-2 was continuing to mutate as expected.

We found no significant difference in the number of IRs in the complete genome or in the *S* genes between variants (Table S1; Figure S1). Unexpectedly, we did find that the number of defining spike mutations occurring within IRs was largely varied between variants (Table 1; Figure 1A). Defining spike mutations of the Delta (22.2%), 20B (25%), and Iota (33.3%) variants were least likely to be found within IRs, but spike mutations of the Beta (70%), Eta (77.7%), and Theta (71.4%) variants were frequently located within IRs (Table 1; Figure 1A). Regarding specific mutations, the D614, N501, ΔY144, ΔG142, T478, N440, K417, Q498, and ΔH69/V70 mutations were most frequently found within IRs (Table 1). Interestingly, we also found that defining mutations shared by the variants (e.g., ΔD69/V70, ΔY144, N501Y, and D614G) were significantly more likely to be found within IRs compared to those unique to a variant, such as A570D, T716I, and S982A in the Alpha variant and ΔE156/F157 and R158G in the Delta variant (Figure 1C,D). We also observed a preference for the defining mutation to be found within the stem rather than the loop of the IR (Figure 1B; Table S1). Thus, it appeared that IRs play an integral role in driving the mutational diversity of spike protein mutations amongst variants. However, why some mutations were preferentially found within IRs and not others was unknown.

Table 1. SARS-CoV-2 variants studied and defining spike mutations containing inverted repeats (IRs).

Variant	Pango Lineage	Defining Spike Protein Mutations	Defining Mutations within IRs
Epsilon	B.1.427	S13I W152C L452R D614G	W152C D614G
20A	B.1.620	ΔP26, ΔH69/V70, V126A, ΔY144, ΔL241-A243, H245Y, S477N, E484K, D614G, P681H, T1027I, D1118H	ΔH69/V70, V126A, ΔY144, ΔL241-A243, D614G, D1118H
Beta	B.1.351	D80A, D215G, ΔL241-A243, K417N, E484K, N501Y, D614G, A701V	ΔL241-A243, K417N, N501Y, D614G, A701V
Alpha	B.1.1.7	ΔH69/V70, ΔY144, N501Y, A570D, D614G, P681H, T716I, S982A, D1118H	ΔH69/V70, ΔY144, N501Y, D614G, D1118H

Table 1. *Cont.*

Variant	Pango Lineage	Defining Spike Protein Mutations	Defining Mutations within IRs
Delta	B.1.617.2	T19R, ΔE156/F157, R158G, L452R, T478K, D614G, P681R, D950N	D614G, D950N
Kappa	B.1.617.1	E154K, L452R, E484Q, D614G, P681R, Q1071H	E484Q, D614G, Q1071H
Gamma	P.1	L18F, T20N, P26S, D138Y, R190S, K417T, E484K, N501Y, D614G, H655Y, T1027I, V1176F	L18F, T20N, R190S, K417T, N501Y, D614G, V1176F
Iota	B.1.526	L5F, T95I, D253G, E484K, D614G, A701V	D614G, A701V
20B	B.1.1.519	T478K, D614G, P681R, T732A	D614G
Eta	B.1.525	Q52R, A67V, ΔH69/V70, ΔY144, E484K, D614G, Q677H, F888L	Q52R, A67V, ΔH69/V70, ΔY144, D614G, F888L
Lambda	C.37	G75V, T76I, ΔR246-G252, D253N, L452Q, F490S, D614G, T859N	G75V, T76I, ΔR246-G252, L452Q, F490S, D614G
Mu	B.1.621	T95I, Y144S, Y145N, R346K, E484K, N501Y, D614G, P681R, D950N	Y144S, Y145N, D614G, D950N
Theta	P.3	E484K, N501Y, D614G, P681R, E1092K, H1101Y, V1176F	N501Y, D614G, E1092K, H1101Y, V1176F
Omicron	BA.1	A67V, ΔH69/V70, T95I, ΔG142-Y144, Y145D, ΔN211, L212I, G339D, S371L, S373P, K417N, N440K, G446S, S477N, T478K, E484A, Q493R, Q496R, Q498R, N501Y, Y505H, T547K, D614G, H655Y, N679K, P681H, N764K, D796Y, N856K, Q954H, N969K, L981F	A67V, ΔH69/V70, ΔG142-Y144, Y145D, ΔN211, L212I, N440K, S477N, T478K, Q498R, N501Y, Y505H, D614G, N679K, N764K, D796Y, Q954H, N969K, L981F
Omicron	BA.2	T19I, ΔL24-P26, A27S, G142D, V213G, G339D, S371F, S373P, S375F, T376A, D405N, R408S, K417N, N440K, S477N, T478K, E484A, Q493R, Q498R, N501Y, Y505H, D614G, H655Y, N679K, P681H, N764K, D796Y, Q954H, N969K	G142D, V213G, N440K, S477N, T478K, Q498R, N501Y, N764K, D796Y, Q954H, N969K
Omicron	BA.4/BA.5	T19I, ΔL24-P26, A27S, ΔH69/V70, G142D, V213G, G339D, S371F, S373P, S375F, T376A, D405N, R408S, K417N, N440K, L452R, S477N, T478K, E484A, F486V, Q498R, N501Y, Y505H, D614G, H665Y, N679K, P681H, N764K, D796Y, Q954H, N969K	ΔH69/V70, G142D, V213G, N440K, L452R, S477N, T478K, F486V, Q498R, N501Y, H665Y, N764K, D796Y, Q954H, N969K
Omicron	BA.2.12.1	T19I, ΔL24-P26, A27S, G142D, V213G, G339D, S371F, S373P, S375F, T376A, D405N, R408S, K417N, N440K, L452Q, S477N, T478K, E484A, Q493A, Q498R, N501Y, Y505H, D614G, H655Y, N679K, P681H, S704L, N764K, D796Y, Q954H, N969K	G142D, V213G, D405N, K417N, N440K, L452Q, S477N, T478K, Q498R, N501Y, D614G, N679K, N764K, D796Y, Q954H, N969K
Omicron	BA.2.75	T19I, ΔL24-P26, A27S, G142D, K147E, W152R, F157L, I210V, V213G, G257S, G339H, S371F, S373P, S375F, T376A, D405N, R408S, K417N, N440K, G446S, N460K, S477N, T478K, E484A, R493Q, Q498R, N501Y, Y505H, D614G, H655Y, N679K, P681H, N764K, D796Y, Q954H, N969K	K147E, V213G, D405N, K417N, N440K, N460K, S477N, T478K, R493Q, Q498R, N501Y, D614G, N679K, N764K, D796Y, Q954H, N969K

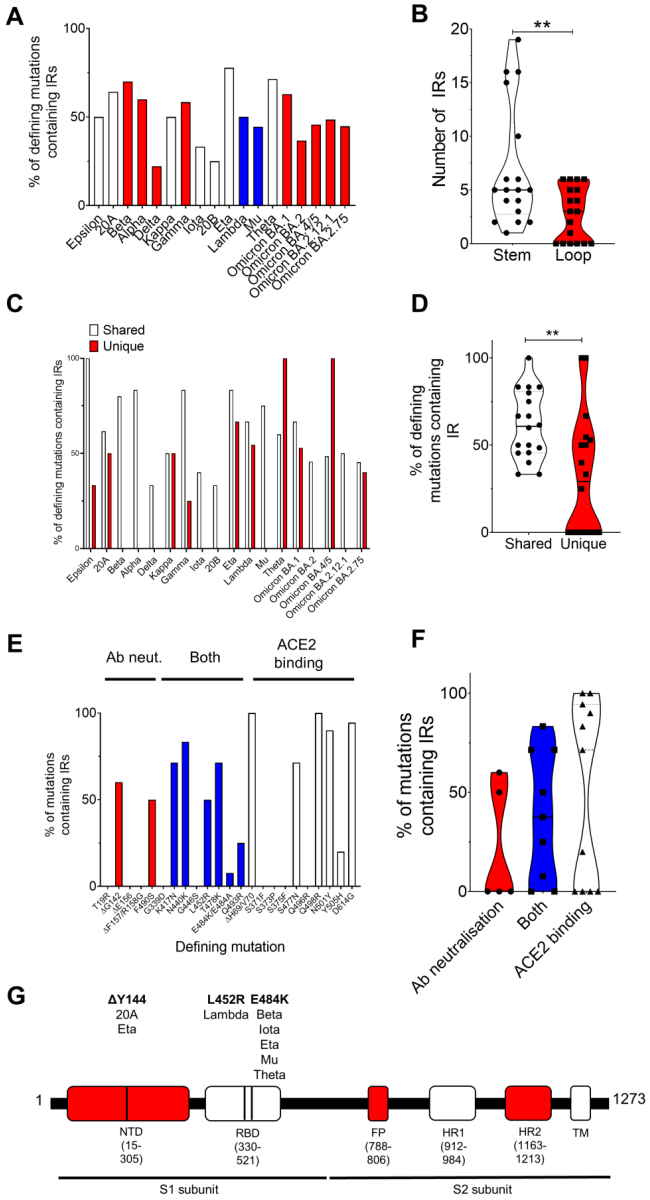

Figure 1. Defining SARS-CoV-2 spike protein mutations involved in ACE2 binding and antibody neutralization are found within inverted repeats (IRs) and predicted pseudoknots. (**A**) The percentage of defining spike protein mutations falling within IRs differed considerably between the variants investigated. (**B**) Spike protein mutations were more likely to fall within the stem, rather than the loop, of the predicted secondary hairpin stem-loop structure. (**C**,**D**) Spike protein mutations which frequently occurred in more than one variant (shared) were significantly more likely to occur within IRs compared to mutations which were unique to a variant. (**E**,**F**) Spike protein mutations associated with ACE2 binding were more frequently found within IRs, whereas mutations associated with antibody neutralization were found infrequently within IRs. Mutations implicated in both are highlighted in blue. (**G**) In the Wuhan strain, pseudoknots were predicted to form within, or within 20 bp, of the ΔY144, L452K, and E484K mutations involved in antibody neutralization and ACE2 binding within the NTD and RBD of the SARS-CoV-2 spike protein. Those variants sharing these predicted pseudoknots are listed. IRs were detected using Palindrome Analyser with the default settings (size: 6–30 bp, spacer: 0–10 bp, and mismatches: 0, 1) and pseudoknot predictions were performed using ProbKnot in the RNAstructure program. ** indicates a significant difference ($p < 0.05$). Data were tested for normality via Shapiro–Wilk prior to comparison via an unpaired Student's *t*-test. NTD, N-terminal domain; RBD, receptor-binding domain; FP, fusion peptide; HR, heptapeptide repeat sequence; TM, transmembrane domain.

Many of the mutations under investigation have now been implicated in ACE2 binding, antibody neutralization, or both [29–34]. We observed that defining mutations contributing to ACE2 binding, such as ΔH69/V70, N501Y, and D614G, were regularly found to occur within IRs (100.0%, 90.0%, and 94.4% of instances where this mutation was present were found within IRs, respectively). Conversely, mutations significantly contributing to antibody neutralization, such as T19R, ΔE156, ΔF157, R158G, and G446S, were not found within IRs (Figure 1E,F; Table S2).

3.2. Pseudoknots Are Predicted to Occur near the Sites of Several Key Mutations

In the Wuhan reference strain, pseudoknot prediction algorithms determined the presence of potential pseudoknots within the sites where the A27S, E484K, and S704L mutations occur (Figure 1G). Moreover, pseudoknots were predicted to form within 30 bp of the sites where T20N, ΔL24-P26, Q52R, G75V, T76I, D80A, R158G, R190S, G257S, R346K, S371L, S373P, S375F, T376A, K417N, G446S, L452R, S477N, T478K, F486V, F490S, A570D, D614G, Q677H, A701V, D796Y, F888L, D950N, and Q954H arise in other variants (Supplementary genome information.

3.3. G4 Are Predicted to Form on the Negative Strand Genome in SARS-CoV-2

We analysed SARS-CoV-2 genomes from all variants and identified putative G4-forming sequences in the *nsp1*, *nsp2*, *nsp3*, *nsp4*, *nsp10*, *nsp12*, *nsp14*, *nsp15*, *nsp16*, *S*, and *N* genes and both the 3′ and 5′ untranslated regions (UTRs; Table 2 and Table S3). Notably, the predicted conserved G4-forming sequence (**GG**AA**GGG**UCCAUUGUUU**GG**UU**GG**; reverse complemented sequence) falls within stem-loop 1 region (SL1; Table 2; Figure 2A). The G4 in the 3′UTR is predicted to fall with the stem-loop II-like motif (S2M), a motif conserved amongst positive single-stranded RNA viruses from the *Astroviridae*, *Calciviridae*, *Picornaviridae*, and *Coronaviridae* (Table 2; Figure 2B) [35]. Both the predicted UTR sequences are the reverse complement and the G4 would be found on the negative sense strand. Interestingly, the Omicron variants BA.2, BA.4, BA.5, BA.2.12.1, and BA.2.75 appear to have lost two of the predicted G4-forming sequences in other variants. These were found in *nsp2* and the 3′UTR S2M motif.

Table 2. SARS-CoV-2 genes containing predicted G-quadruplex (G4)-forming sequences.

Gene Name/Region	Highest Scoring Sequence	+ or — Strand
nsp1	**GG**CTTT**GG**A**G**ACTCC**G**T**GG**A**GG**A**GG**	+
nsp2	**GG**T**G**TT**G**TT**GG**A**G**AA**GG**TTCC**G**AA**GG**	+
nsp3	**GG**ATAT**GG**TT**GG**TTT**GG**	–
nsp4	**GG**T**G**ATA**G**A**GG**TTT**G**T**GG**T**GG**TT**GG**	–
nsp10	**GG**TAT**G**T**GG**AAA**GG**TTAT**GG**	+
nsp12	**GG**AACCACTAAATTTTAT**GG**T**GG**TT**GG**	+
nsp14	**GG**TT**GG**TT**GG**TTTT**G**AT**G**TT**G**AA**GG**	+
nsp15	**GG**A**G**CCCACAA**GG**TAATCCA**GG**T**GG**	+
nsp16	**GG**A**G**AAATA**G**TACAACAT**GG**AAT**GG**C**GG**	+
S	**GG**CTTATA**GG**TTTAAT**GG**TATT**GG**	+
N	**GG**CT**GG**CAAT**GG**C**GG**	+
3′UTR	**GG**U**GG**U**G**TAAAA**G**U**GG**CUCC**GG**	–
5′UTR	**GG**AA**GGG**UCCAUU**G**UUU**GG**UU**GG**	–

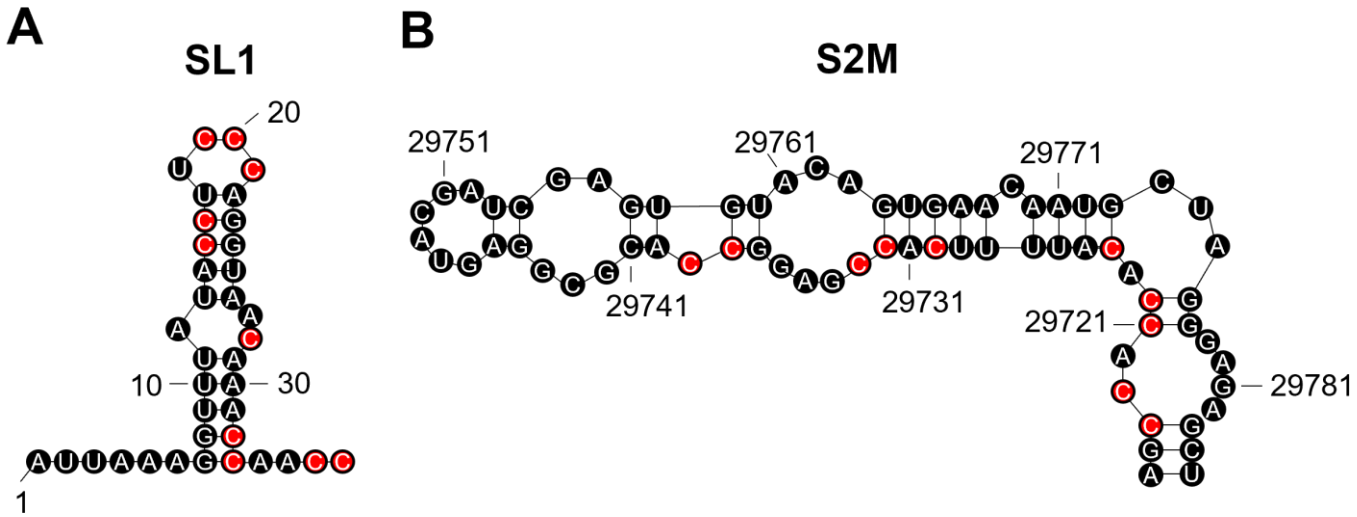

Figure 2. The stem-loop 1 (SL1) and stem-loop II-like motif (S2M) regions in the SARS-CoV-2 untranslated regions (UTRs) contain sequences on their positive strand which may allow G-quadruplex (G4) formation on the negative strand. Both the SL1 (**A**) and S2M (**B**) regions contain C-tracts on their positive sense RNA strand which may allow the formation of a G4 to occur on the negative sense strand. The cytosines which may be involved are highlighted in red. The reverse complemented sequence on the negative strand would be GGAAGGGUCCAUUGUUUGGUUGG (SL1) and GGUGGUGTAAAAGUGGCUCCGG (S2M). Both sequences have the potential to form G4s and G4-formation on the negative strand may prevent viral translation and replication.

4. Discussion

We found that there was significant diversity in the percentage of defining spike protein mutations occurring within IRs between variants. Mutations linked to infectivity were more likely to arise within IRs compared to those associated with antibody neutralization. Moreover, pseudoknots were predicted to form close to key spike protein mutations and G4s were predicted to form within two conserved regions within the 3′ and 5′ UTRs.

The SARS-CoV-2 mutations found most likely to occur within IRs amongst all variants were the ΔH69/V70, N440K, N501Y, and D614G mutations; all of which have been implicated in increased fitness and infectivity [36–40]. Mutations implicated in ACE2 binding and propagation were found to frequently occur within IRs, whilst mutations involved in antibody neutralization and reduced vaccine efficacy were rarely found within IRs. Although not significant, these data suggest that mutations linked to antibody neutralization may occur more frequently outside of IRs and are probably evoked due to the external pressures of vaccines and antibodies, rather than spontaneous mutation.

Surprisingly, the D614G mutation was found within an IR for all variants except for Omicron BA.2, BA.4, and BA.5. The D614G mutation has been shown to enhance infectivity but it has also been shown to enhance susceptibility to vaccines and antibody neutralization. Notably, the Omicron BA.2, BA.4, and BA.5 variants also display increased resistance to neutralizing antibodies [41–44]. However, this is unlikely to be due to the loss of this IR sequence and more likely to be due to the involvement of the S371F, D405N, R408S, F486V, and L452R mutations. However, of these mutations, only the F486V and L452R mutations were found within IRs, further supporting our claim that antibody neutralizing mutations occur with less frequency within IRs.

In the Wuhan reference strain, pseudoknot prediction algorithms determined the presence of potential pseudoknots within the sites where the A27S, E484K, and S704L mutations occur and in 29 additional mutations amongst the variants tested. The E484 mutation is particularly noteworthy as this mutation has been shown to arise with high frequency in the presence of antibodies [45]. However, whether the external influence exerted by antibodies can induce pseudoknot formation is unknown. It is well known that

the pseudoknot in the ORF1 polyprotein of SARS-CoV-2 can induce frameshifts, whilst the conserved pseudoknot in the coronavirus 3′UTR is involved in viral replication [46]. These are two such key examples, but one can hypothesise that these examples are the tip of the iceberg, and that pseudoknots have important roles throughout the entire viral genome. Future studies could investigate whether antibody binding acts as an environmental trigger for pseudoknot formation/prevention and whether this influences further mutational drive.

It was previously demonstrated that the SARS-CoV-2 genome contained fewer G4s than the SARS-CoV genome and this has been suggested to be energetically favourable, as G4s can represent a barrier to translation and replication [47,48]. Moreover, the frequency of G4s in a viral genome is associated with whether infection is chronic or acute [28]. The G4s in the *nsp1*, *nsp3*, *nsp10*, *S*, and *N* genes have previously been shown to form in vitro [48]. However, the roles these G4s might play in controlling the biological functions of these genes have not been fully addressed. Of particular interest are the G4-forming sequences found in the UTRs of SARS-CoV-2. Both predicted G4 sequences would be found on the negative strand. It has recently been identified that SARS-CoV-2 negative strands have protein-coding potential, and they are known to be involved in replication [49]. Thus, the negative strand may be targeted by G4-stabilising compounds to prevent translation of proteins on the negative sense strand and subsequent SARS-CoV-2 replication cycles. Indeed, several G4-stabilisers have been found to bind to SARS-CoV-2 RNA and G4-stabilising compounds have recently been demonstrated to be antiviral in mouse models of infection [50–52], highlighting the therapeutic potential of targeting G4s in SARS-CoV-2 infections.

It has recently been shown that the conserved SL1 region in the 5′UTR of SARS-CoV-2 represents a potential drug target [53]. The authors demonstrated that a locked nucleic acid (LNA) antisense oligonucleotide to the SL1 region could inhibit viral translation, prevent lethality in mice expressing ACE2, and make SARS-CoV-2 vulnerable to non-structural protein 1 (Nsp1) translation suppression [53]. Chowdhury et al., recently demonstrated that LNA probes can promote disruption of the secondary G4 structure [54]. Therefore, it is likely that the LNA oligonucleotide used against the SL1 region could also disrupt the G4 predicted to form on the negative strand. This suggests that this conserved G4-forming sequence could be important in promoting viral translation and molecules designed to disrupt this G4 might have therapeutic potential.

The S2M region has previously been described as a recombination hotspot in SARS-CoV-2 compared with other positive single-stranded RNA viruses [55]. It is well-established that G4s can contribute to genome instability, and it is likely that this G4-forming sequence in past variants has contributed to the genetic variability observed within the S2M region of the new variants. On another note, the presence of two potential TAGGGA microsatellites in close vicinity to this region probably also contributes to the genetic variability within this region due to their high mutation rates. Finally, interferon-β (IFN-β) can inhibit SARS-CoV-2 replication and Nsp2 has recently been shown to repress the translation of IFN-β [56]. The presence of G4-forming sequences in an mRNA can prevent translation, and the loss of the predicted sequences in the recent Omicron variants could provide some explanation for the increased replication of these variants. Thus, loss of the G4-forming sequence from *nsp2* might enhance the translation of Nsp2 and promote replication.

5. Conclusions

Taken together, non-B nucleic acids structures are prevalent throughout the SARS-CoV-2 genome where they may play integral roles in promoting mutational diversity. Furthermore, it could be interesting to explore whether environmental pressures, such as the immune response and antibodies, influence the formation of IRs, G4s, and pseudoknots. Finally, targeting non-B nucleic acids structures in SARS-CoV-2 may disrupt viral biological processes and have therapeutic potential, although a much greater understanding of their biological roles in SARS-CoV-2 is required.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/genes14010157/s1, Figure S1: The total number of IRs in the whole genome and spike protein; Table S1: SARS-CoV-2 spike protein mutations within inverted repeats (IRs); Table S2: Mutations involved in ACE2 binding and antibody neutralization found within IRs; and Table S3: Putative G4-forming sequences in SARS-CoV-2; Supplementary genome information.

Author Contributions: Conceptualization, S.B. and V.B.; Formal analysis, S.B.; Funding acquisition, V.B.; Investigation, S.B. and V.B.; Methodology, S.B. and V.B.; Project administration, S.B.; Visualization, S.B.; Writing—original draft, S.B.; Writing—review and editing, S.B. and V.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Czech Science Foundation, grant number 22-21903S.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data supporting the study are available upon request.

Acknowledgments: The authors would like to thank those who are continually working hard to sequence SARS-CoV-2 genomes; their endeavours have been invaluable in our ongoing fight against SARS-CoV-2.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Skourti-Stathaki, K.; Proudfoot, N.J. A Double-Edged Sword: R Loops as Threats to Genome Integrity and Powerful Regulators of Gene Expression. *Genes Dev.* **2014**, *28*, 1384–1396. [CrossRef] [PubMed]
2. Voineagu, I.; Narayanan, V.; Lobachev, K.S.; Mirkin, S.M. Replication Stalling at Unstable Inverted Repeats: Interplay between DNA Hairpins and Fork Stabilizing Proteins. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 9936–9941. [CrossRef] [PubMed]
3. Saranathan, N.; Vivekanandan, P. G-Quadruplexes: More Than Just a Kink in Microbial Genomes. *Trends Microbiol.* **2019**, *27*, 148–163. [CrossRef]
4. Griffin, B.D.; Bass, H.W. Review: Plant G-Quadruplex (G4) Motifs in DNA and RNA; Abundant, Intriguing Sequences of Unknown Function. *Plant Sci.* **2018**, *269*, 143–147. [CrossRef]
5. Brierley, I.; Pennell, S.; Gilbert, R.J.C. Viral RNA Pseudoknots: Versatile Motifs in Gene Expression and Replication. *Nat. Rev. Microbiol.* **2007**, *5*, 598–610. [CrossRef]
6. Pearson, C.E.; Zorbas, H.; Price, G.B.; Zannis-Hadjopoulos, M. Inverted Repeats, Stem-Loops, and Cruciforms: Significance for Initiation of DNA Replication. *J. Cell. Biochem.* **1996**, *63*, 1–22. [CrossRef]
7. del Solar, G.; Giraldo, R.; Ruiz-Echevarría, M.J.; Espinosa, M.; Díaz-Orejas, R. Replication and Control of Circular Bacterial Plasmids. *Microbiol. Mol. Biol. Rev.* **1998**, *62*, 434–464. [CrossRef] [PubMed]
8. Lu, S.; Wang, G.; Bacolla, A.; Zhao, J.; Spitser, S.; Vasquez, K.M. Short Inverted Repeats Are Hotspots for Genetic Instability: Relevance to Cancer Genomes. *Cell Rep.* **2015**, *10*, 1674–1680. [CrossRef]
9. Sadler, J.R.; Sasmor, H.; Betz, J.L. A Perfectly Symmetric Lac Operator Binds the Lac Repressor Very Tightly. *Proc. Natl. Acad. Sci. USA* **1983**, *80*, 6785–6789. [CrossRef]
10. Butler, D.K.; Yasuda, L.E.; Yao, M.C. Induction of Large DNA Palindrome Formation in Yeast: Implications for Gene Amplification and Genome Stability in Eukaryotes. *Cell* **1996**, *87*, 1115–1122. [CrossRef]
11. Okamura, K.; Chung, W.-J.; Lai, E.C. The Long and Short of Inverted Repeat Genes in Animals: MicroRNAs, Mirtrons and Hairpin RNAs. *Cell Cycle* **2008**, *7*, 2840–2845. [CrossRef] [PubMed]
12. Wroblewski, T.; Matvienko, M.; Piskurewicz, U.; Xu, H.; Martineau, B.; Wong, J.; Govindarajulu, M.; Kozik, A.; Michelmore, R.W. Distinctive Profiles of Small RNA Couple Inverted Repeat-Induced Post-Transcriptional Gene Silencing with Endogenous RNA Silencing Pathways in Arabidopsis. *RNA* **2014**, *20*, 1987–1999. [CrossRef] [PubMed]
13. Staple, D.W.; Butcher, S.E. Pseudoknots: RNA Structures with Diverse Functions. *PLoS Biol.* **2005**, *3*, e213. [CrossRef] [PubMed]
14. Neupane, K.; Munshi, S.; Zhao, M.; Ritchie, D.B.; Ileperuma, S.M.; Woodside, M.T. Anti-Frameshifting Ligand Active against SARS Coronavirus-2 Is Resistant to Natural Mutations of the Frameshift-Stimulatory Pseudoknot. *J. Mol. Biol.* **2020**, *432*, 5843–5847. [CrossRef] [PubMed]
15. Varshney, D.; Spiegel, J.; Zyner, K.; Tannahill, D.; Balasubramanian, S. The Regulation and Functions of DNA and RNA G-Quadruplexes. *Nat. Rev. Mol. Cell Biol.* **2020**, *21*, 459–474. [CrossRef]
16. Cebrián, R.; Belmonte-Reche, E.; Pirota, V.; de Jong, A.; Morales, J.C.; Freccero, M.; Doria, F.; Kuipers, O.P. G-Quadruplex DNA as a Target in Pathogenic Bacteria: Efficacy of an Extended Naphthalene Diimide Ligand and Its Mode of Action. *J. Med. Chem.* **2022**, *65*, 4752–4766. [CrossRef] [PubMed]

17. Abiri, A.; Lavigne, M.; Rezaei, M.; Nikzad, S.; Zare, P.; Mergny, J.-L.; Rahimi, H.-R. Unlocking G-Quadruplexes as Antiviral Targets. *Pharmacol. Rev.* **2021**, *73*, 897–923. [CrossRef]
18. Cantara, A.; Luo, Y.; Dobrovolná, M.; Bohalova, N.; Fojta, M.; Verga, D.; Guittat, L.; Cucchiarini, A.; Savrimoutou, S.; Häberli, C.; et al. G-Quadruplexes in Helminth Parasites. *Nucleic Acids Res.* **2022**, *50*, 2719–2735. [CrossRef]
19. Warner, E.F.; Bohálová, N.; Brázda, V.; Waller, Z.A.E.; Bidula, S. Analysis of Putative Quadruplex-Forming Sequences in Fungal Genomes: Novel Antifungal Targets? *Microb. Genom.* **2021**, *7*, 000570. [CrossRef]
20. Goswami, P.; Bartas, M.; Lexa, M.; Bohálová, N.; Volná, A.; Červeň, J.; Červeňová, V.; Pečinka, P.; Špunda, V.; Fojta, M.; et al. SARS-CoV-2 Hot-Spot Mutations Are Significantly Enriched within Inverted Repeats and CpG Island Loci. *Brief. Bioinform.* **2021**, *22*, 1338–1345. [CrossRef]
21. Bartas, M.; Goswami, P.; Lexa, M.; Červeň, J.; Volná, A.; Fojta, M.; Brázda, V.; Pečinka, P. Letter to the Editor: Significant Mutation Enrichment in Inverted Repeat Sites of New SARS-CoV-2 Strains. *Brief. Bioinform.* **2021**, *22*, bbab129. [CrossRef] [PubMed]
22. Zhao, C.; Qin, G.; Niu, J.; Wang, Z.; Wang, C.; Ren, J.; Qu, X. Targeting RNA G-Quadruplex in SARS-CoV-2: A Promising Therapeutic Target for COVID-19? *Angew. Chem. Int. Ed.* **2021**, *60*, 432–438. [CrossRef] [PubMed]
23. Brázda, V.; Kolomazník, J.; Lýsek, J.; Hároníková, L.; Coufal, J.; Štastný, J. Palindrome Analyser—A New Web-Based Server for Predicting and Evaluating Inverted Repeats in Nucleotide Sequences. *Biochem. Biophys. Res. Commun.* **2016**, *478*, 1739–1745. [CrossRef] [PubMed]
24. Hodcroft, E.B. CoVariants: SARS-CoV-2 Mutations and Variants of Interest. Available online: https://covariants.org/ (accessed on 19 December 2022).
25. Shu, Y.; McCauley, J. GISAID: Global Initiative on Sharing All Influenza Data—From Vision to Reality. *Euro Surveill.* **2017**, *22*, 30494. [CrossRef] [PubMed]
26. Smit, S.; Rother, K.; Heringa, J.; Knight, R. From Knotted to Nested RNA Structures: A Variety of Computational Methods for Pseudoknot Removal. *RNA* **2008**, *14*, 410–416. [CrossRef] [PubMed]
27. Bellaousov, S.; Mathews, D.H. ProbKnot: Fast Prediction of RNA Secondary Structure Including Pseudoknots. *RNA* **2010**, *16*, 1870–1880. [CrossRef]
28. Kikin, O.; D'Antonio, L.; Bagga, P.S. QGRS Mapper: A Web-Based Server for Predicting G-Quadruplexes in Nucleotide Sequences. *Nucleic Acids Res.* **2006**, *34*, W676–W682. [CrossRef]
29. Yang, T.-J.; Yu, P.-Y.; Chang, Y.-C.; Liang, K.-H.; Tso, H.-C.; Ho, M.-R.; Chen, W.-Y.; Lin, H.-T.; Wu, H.-C.; Hsu, S.-T.D. Effect of SARS-CoV-2 B.1.1.7 Mutations on Spike Protein Structure and Function. *Nat. Struct. Mol. Biol.* **2021**, *28*, 731–739. [CrossRef]
30. Hirabara, S.M.; Serdan, T.D.A.; Gorjao, R.; Masi, L.N.; Pithon-Curi, T.C.; Covas, D.T.; Curi, R.; Durigon, E.L. SARS-CoV-2 Variants: Differences and Potential of Immune Evasion. *Front. Cell. Infect. Microbiol.* **2021**, *11*, 781429. [CrossRef]
31. Harvey, W.T.; Carabelli, A.M.; Jackson, B.; Gupta, R.K.; Thomson, E.C.; Harrison, E.M.; Ludden, C.; Reeve, R.; Rambaut, A.; Peacock, S.J.; et al. SARS-CoV-2 Variants, Spike Mutations and Immune Escape. *Nat. Rev. Microbiol.* **2021**, *19*, 409–424. [CrossRef]
32. McCallum, M.; Walls, A.C.; Sprouse, K.R.; Bowen, J.E.; Rosen, L.E.; Dang, H.V.; De Marco, A.; Franko, N.; Tilles, S.W.; Logue, J.; et al. Molecular Basis of Immune Evasion by the Delta and Kappa SARS-CoV-2 Variants. *Science* **2021**, *374*, 1621–1626. [CrossRef] [PubMed]
33. Chen, R.E.; Zhang, X.; Case, J.B.; Winkler, E.S.; Liu, Y.; VanBlargan, L.A.; Liu, J.; Errico, J.M.; Xie, X.; Suryadevara, N.; et al. Resistance of SARS-CoV-2 Variants to Neutralization by Monoclonal and Serum-Derived Polyclonal Antibodies. *Nat. Med.* **2021**, *27*, 717–726. [CrossRef] [PubMed]
34. Barton, M.I.; MacGowan, S.A.; Kutuzov, M.A.; Dushek, O.; Barton, G.J.; van der Merwe, P.A. Effects of Common Mutations in the SARS-CoV-2 Spike RBD and Its Ligand, the Human ACE2 Receptor on Binding Affinity and Kinetics. *Elife* **2021**, *10*, e70658. [CrossRef] [PubMed]
35. Imperatore, J.A.; Cunningham, C.L.; Pellegrene, K.A.; Brinson, R.G.; Marino, J.P.; Evanseck, J.D.; Mihailescu, M.R. Highly Conserved S2m Element of SARS-CoV-2 Dimerizes via a Kissing Complex and Interacts with Host MiRNA-1307-3p. *Nucleic Acids Res.* **2022**, *50*, 1017–1032. [CrossRef] [PubMed]
36. Tandel, D.; Gupta, D.; Sah, V.; Harshan, K.H. N440K Variant of SARS-CoV-2 Has Higher Infectious Fitness. *bioRxiv* **2021**. [CrossRef]
37. Bate, N.; Savva, C.G.; Moody, P.C.E.; Brown, E.A.; Evans, S.E.; Ball, J.K.; Schwabe, J.W.R.; Sale, J.E.; Brindle, N.P.J. In Vitro Evolution Predicts Emerging SARS-CoV-2 Mutations with High Affinity for ACE2 and Cross-Species Binding. *PLoS Pathog.* **2022**, *18*, e1010733. [CrossRef]
38. Korber, B.; Fischer, W.M.; Gnanakaran, S.; Yoon, H.; Theiler, J.; Abfalterer, W.; Hengartner, N.; Giorgi, E.E.; Bhattacharya, T.; Foley, B.; et al. Tracking Changes in SARS-CoV-2 Spike: Evidence That D614G Increases Infectivity of the COVID-19 Virus. *Cell* **2020**, *182*, 812–827.e19. [CrossRef]
39. Tian, F.; Tong, B.; Sun, L.; Shi, S.; Zheng, B.; Wang, Z.; Dong, X.; Zheng, P. N501Y Mutation of Spike Protein in SARS-CoV-2 Strengthens Its Binding to Receptor ACE2. *Elife* **2021**, *10*, e69091. [CrossRef]
40. Meng, B.; Kemp, S.A.; Papa, G.; Datir, R.; Ferreira, I.A.T.M.; Marelli, S.; Harvey, W.T.; Lytras, S.; Mohamed, A.; Gallo, G.; et al. Recurrent Emergence of SARS-CoV-2 Spike Deletion H69/V70 and Its Role in the Alpha Variant B.1.1.7. *Cell Rep.* **2021**, *35*, 109292. [CrossRef]
41. Iketani, S.; Liu, L.; Guo, Y.; Liu, L.; Chan, J.F.-W.; Huang, Y.; Wang, M.; Luo, Y.; Yu, J.; Chu, H.; et al. Antibody Evasion Properties of SARS-CoV-2 Omicron Sublineages. *Nature* **2022**, *604*, 553–556. [CrossRef]

42. Zhou, H.; Dcosta, B.M.; Landau, N.R.; Tada, T. Resistance of SARS-CoV-2 Omicron BA.1 and BA.2 Variants to Vaccine-Elicited Sera and Therapeutic Monoclonal Antibodies. *Viruses* **2022**, *14*, 1334. [CrossRef] [PubMed]
43. Yamasoba, D.; Kosugi, Y.; Kimura, I.; Fujita, S.; Uriu, K.; Ito, J.; Sato, K. Neutralisation Sensitivity of SARS-CoV-2 Omicron Subvariants to Therapeutic Monoclonal Antibodies. *Lancet Infect. Dis.* **2022**, *22*, 942–943. [CrossRef] [PubMed]
44. Cao, Y.; Yisimayi, A.; Jian, F.; Song, W.; Xiao, T.; Wang, L.; Du, S.; Wang, J.; Li, Q.; Chen, X.; et al. BA.2.12.1, BA.4 and BA.5 Escape Antibodies Elicited by Omicron Infection. *Nature* **2022**, *608*, 593–602. [CrossRef]
45. Weisblum, Y.; Schmidt, F.; Zhang, F.; DaSilva, J.; Poston, D.; Lorenzi, J.C.; Muecksch, F.; Rutkowska, M.; Hoffmann, H.-H.; Michailidis, E.; et al. Escape from Neutralizing Antibodies by SARS-CoV-2 Spike Protein Variants. *Elife* **2020**, *9*, e61312. [CrossRef] [PubMed]
46. Williams, G.D.; Chang, R.Y.; Brian, D.A. A Phylogenetically Conserved Hairpin-Type 3′ Untranslated Region Pseudoknot Functions in Coronavirus RNA Replication. *J. Virol.* **1999**, *73*, 8349–8355. [CrossRef] [PubMed]
47. Belmonte-Reche, E.; Serrano-Chacón, I.; Gonzalez, C.; Gallo, J.; Bañobre-López, M. Potential G-Quadruplexes and i-Motifs in the SARS-CoV-2. *PLoS ONE* **2021**, *16*, e0250654. [CrossRef]
48. Cui, H.; Zhang, L. G-Quadruplexes Are Present in Human Coronaviruses Including SARS-CoV-2. *Front. Microbiol.* **2020**, *11*, 567317. [CrossRef] [PubMed]
49. Dinan, A.M.; Lukhovitskaya, N.I.; Olendraite, I.; Firth, A.E. A Case for a Negative-Strand Coding Sequence in a Group of Positive-Sense RNA Viruses. *Virus Evol.* **2020**, *6*, veaa007. [CrossRef]
50. Liu, G.; Du, W.; Sang, X.; Tong, Q.; Wang, Y.; Chen, G.; Yuan, Y.; Jiang, L.; Cheng, W.; Liu, D.; et al. RNA G-Quadruplex in TMPRSS2 Reduces SARS-CoV-2 Infection. *Nat. Commun.* **2022**, *13*, 1444. [CrossRef] [PubMed]
51. Moraca, F.; Marzano, S.; D'Amico, F.; Lupia, A.; Di Fonzo, S.; Vertecchi, E.; Salvati, E.; Di Porzio, A.; Catalanotti, B.; Randazzo, A.; et al. Ligand-Based Drug Repurposing Strategy Identified SARS-CoV-2 RNA G-Quadruplex Binders. *Chem. Commun.* **2022**, *58*, 11913–11916. [CrossRef] [PubMed]
52. Qin, G.; Zhao, C.; Liu, Y.; Zhang, C.; Yang, G.; Yang, J.; Wang, Z.; Wang, C.; Tu, C.; Guo, Z.; et al. RNA G-Quadruplex Formed in SARS-CoV-2 Used for COVID-19 Treatment in Animal Models. *Cell Discov.* **2022**, *8*, 86. [CrossRef] [PubMed]
53. Vora, S.M.; Fontana, P.; Mao, T.; Leger, V.; Zhang, Y.; Fu, T.-M.; Lieberman, J.; Gehrke, L.; Shi, M.; Wang, L.; et al. Targeting Stem-Loop 1 of the SARS-CoV-2 5′ UTR to Suppress Viral Translation and Nsp1 Evasion. *Proc. Natl. Acad. Sci. USA* **2022**, *119*, e2117198119. [CrossRef] [PubMed]
54. Chowdhury, S.; Wang, J.; Nuccio, S.P.; Mao, H.; Di Antonio, M. Short LNA-Modified Oligonucleotide Probes as Efficient Disruptors of DNA G-Quadruplexes. *Nucleic Acids Res.* **2022**, *50*, 7247–7259. [CrossRef] [PubMed]
55. Yeh, T.-Y.; Contreras, G.P. Emerging Viral Mutants in Australia Suggest RNA Recombination Event in the SARS-CoV-2 Genome. *Med. J. Aust.* **2020**, *213*, 44–44.e1. [CrossRef]
56. Xu, Z.; Choi, J.-H.; Dai, D.L.; Luo, J.; Ladak, R.J.; Li, Q.; Wang, Y.; Zhang, C.; Wiebe, S.; Liu, A.C.H.; et al. SARS-CoV-2 Impairs Interferon Production via NSP2-Induced Repression of MRNA Translation. *Proc. Natl. Acad. Sci. USA* **2022**, *119*, e2204539119. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.