Biochemical and Biophysical Research Communications 569 (2021) 154–160

Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

journal homepage: www.elsevier.com/locate/ybbrc

Analysis of SARS-CoV-2 nucleocapsid phosphoprotein N variations in the binding site to human 14-3-3 proteins

Samanta Del Veliz [a, 1], Lautaro Rivera [a, 1], Diego M. Bustos [a, b], Marina Uhart [a, *]

[a] Laboratorio de Integración de Señales Celulares, IHEM, Universidad Nacional de Cuyo, CONICET, Mendoza, Argentina
[b] Facultad de Ciencias Exactas y Naturales, Universidad Nacional de Cuyo, Mendoza, Argentina

ARTICLE INFO

Article history:
Received 25 June 2021
Accepted 30 June 2021
Available online 2 July 2021

Keywords:
Coronavirus
SARS-CoV-2
N protein
14-3-3
Interaction
Viral success

ABSTRACT

The SARS-CoV-2 N protein binds several cell host proteins including 14-3-3γ, a well-characterized regulatory protein. However, the biological function of this interaction is not completely understood. We analyzed the variability of ~90 000 sequences of the SARS-CoV-2 N protein, particularly, its mutations in disordered regions containing binding motifs for 14-3-3 proteins. We studied how these mutations affect the binding energy to 14-3-3γ and found that changes positively affecting the predicted interaction with 14-3-3γ are the most successfully spread, with the highest prevalence in the phylogenetic tree. Although most residues are highly conserved within the 14-3-3 binding site, compensatory mutations to maintain the interaction energy of N-14-3-3γ were found, including half of the current variants of concern and interest. Our results suggest that binding of N to 14-3-3γ is beneficial for the virus, thus targeting this viral-host protein-protein interaction seems an attractive approach to explore antiviral strategies.

Introduction

The 14-3-3s constitute a family of highly conserved eukaryotic proteins with more than 300 known binding partners, and 3000 predicted. Their binding to specific phosphorylated motifs modifies the activity, stability, and/or subcellular localization of the target proteins [1].

14-3-3 proteins contribute positively in the pathogenesis and progression of viral infections. Viral interactions with 14-3-3 proteins may change the original distribution and functions of this host family. For example, some viruses associate with and inhibit 14-3-3s from inducing apoptosis [2].

The Severe Acute Respiratory Syndrome SARS Coronavirus (SARS-CoV) Nucleocapsid (N) protein is frequently phosphorylated on several Ser/Thr residues in host cells [3], where it is mainly cytoplasmic [4]. It binds to nucleic acids, controlling viral transcription, replication, and assembly [5].

Phosphorylated N protein binds to 14-3-3 [4,6], which regulates its subcellular localization. Decreasing the levels of 14-3-3 using siRNA causes N protein nuclear concentration [4].

Sequence mutations in the Serine-Arginine (SR)-rich region (amino acids 180 to 210) were previously analyzed [7]. Tung and Limtung proposed that only the SARS-CoV-2 N protein monomer can bind the 14-3-3 dimer and that the phosphorylation of the N protein and its sequestration by 14-3-3 is a cellular defense mechanism to control the SARS-CoV-2 through hindering its replication, transcription, and packaging [7]. Nevertheless, Sluchanko and coll [6]. proposed that the N-14-3-3 interaction may sequester 14-3-3 proteins and impedes them to exert their cellular functions, suggesting that this interaction could be beneficial for the virus. They demonstrated that the conserved Ser197 in N protein is essential for the binding to the 14-3-3, in particular to 14-3-3γ paralog [6].

Here, we analyzed the variability of ~90 000 SARS-CoV-2 N protein sequences, and studied how these mutations affected the ΔG of binding to 14-3-3γ. We analyzed if the variants bearing those mutations were successfully spread, and their implications in the discussion of the beneficial or detrimental role of N protein-14-3-3 interaction in SARS-CoV-2 life cycle. We compared the data with the consensus sequences of current circulating variants of concern and interest. This could be relevant to the planning of novel antiviral strategies.

* Corresponding author. IHEM, U.N.Cuyo-CONICET, Av. Libertador 80, 5500, Mendoza, Argentina.
E-mail address: muhart@mendoza-conicet.gob.ar (M. Uhart).
[1] Both authors contributed equally.

https://doi.org/10.1016/j.bbrc.2021.06.100
0006-291X/

Methods

Sequences

All SARS-CoV-2 N protein sequences analyzed in the present report (93,246) were downloaded from GISAID [8] (www.gisaid.org) in August 2020. Low quality or incomplete sequences were excluded from the analysis.

Multiple sequence alignment (MSA)

The 86,168 N protein sequences were aligned using the algorithm Clustal Omega v1.2.4 with parameters in their default setting. Five sequences containing amino acid insertions of more than 10 amino acids were removed, leaving 86,163 sequences. Mutation frequencies were computed using the Balcony package in Rstudio, and the score used in Fig. 2 (A, B) was calculated using the BLOSUM62 (**BLO**cks **SU**bstitution **M**atrix) matrix [9]. In our color scale, white represents conservative changes (score 0, as expected by chance) or no change, blue (positive score) indicates that the change was found in the database more, and red (negative score) less often than expected by chance respectively.

Computer Programming, Statistics, Intrinsic Structural Disorder Calculation, and Eukaryotic Linear Motif (ELM) Identification

Scripts for sequence analyses were programmed in bash. All statistical analyses were carried out using the Rstudio. Disorder predictions were done by using the software Espritz [10] (http://protein. bio.unipd.it/espritz/) as previously done [11]. Briefly, the SARS-CoV-2 N protein sequences were identified and translated from the downloaded SARS-CoV-2 genomes, and then Espritz was run on the sequences, resulting files were post-processed using bash scripting. ELMs for the binding of 14-3-3 proteins on the reference sequence of SARS-CoV-2 N protein were identified using the server http://elm.eu.org/

Entropy calculation

To compare the variability of each position in the MSA we scored the position entropy based on Shannon entropy (S), given by:

$$S_{(i)} = -\sum_{x} p_x(i) \log_2 p_x(i)$$

where i is the MSA interest position, x is any amino acid and $p_{x\,is}$ the frequency of x at i. The calculations were done in R using the package *bios2cor* [12] and plots were done with the package *ggplot2* [13].

Protein-protein affinity calculation

The effects of single-point mutations on 14-3-3γ/peptide interactions were evaluated in the following 3D structures from the RBSD PDB (2B05, 4J6S, 4O46, 5D3E, 6FEL, 6GKF, 6GKG, 6S9K, 6SAD) using the mcsm-ppi2 algorithm [14] and Foldx [15]. Each 3D structure corresponded to 14-3-3γ in complex to a phosphorylated peptide. The point mutations of the amino acids in the crystal to the corresponding amino acids in SARS-CoV-2 N protein were done using the commands PositionScan, Optimize, and AnalyseComplex in FoldX or by using the PDB file and a list of point mutations in mcsm-ppi2.

Amino acid Co-evolution

SARS-CoV-2 N protein amino acid co-variation was evaluated by different methods, Observed Minus Expected Squared (OMES [12],), which is based on the $\chi 2$ test and the Direct Coupling Analysis (DCA [16],) that infers co-evolution between pairs of residues that are spatially close.

The OMES score was computed as:

$$OMES(i,j) = \frac{1}{N_{(i,j)}} \sum_{x,y} \left(N^{obs}_{x,y}(i,j) - N^{ex}_{x,y}(i,j)^2 \right)$$

where N (i, j) is the number of aligned sequences without gaps, at positions i and j, $N^{obs}{}_{xy}$ (i,j) and $N^{ex}{}_{xy}$ (i,j) are the number of times each pair of amino acids (x, y) is observed and expected at (i,j), respectively.

The DCA score was computed as:

$$H_{(DCXA)}(S/J,h) = -\sum_{i<j}^{1,L} J_{ij}(S_i,S_j) - \sum_{i=j}^{L} h_i(S_i)$$

The sequence S = (S1, …, SL) is aligned to the MSA of length L; J and h are known parameters from a seed alignment. The energy $H_{(DCA)}$ is considered as a "score" (lower energy corresponds to higher score) for sequence S.

Results

The host expressed SARS-CoV-2 N protein contains multiple phosphorylation sites [6]. Its most phosphorylated portion is the SR-rich region, located downstream to the N-terminal domain (NTD, Fig. 1A). Based on reports about the earlier SARS-CoV-1 N protein interaction with host 14-3-3 proteins [4], and recent *in silico* data suggesting the interaction has been conserved in SARS-CoV-2 [7], we were interested in studying 14-3-3 interaction with SARS-CoV-2 N protein and the amino acidic mutations occurring in its 14-3-3 binding site/s. Because we have previously shown that the intrinsic disorder is crucial for the binding to 14-3-3s [17], we first analyzed the disorder probability in a SARS-CoV-2 reference amino acidic sequence (EPI_ISL_412977, Wuhan, China, Fig. 1B). We identified 4 disordered regions, one in the N-terminal domain (NTD), two nearby regions linking the NTD and the C-terminal Domain (CTD), and one in the CTD (Fig. 1B). We downloaded 93,246 SARS-CoV-2 N protein sequences available from GISAID [8] (www.gisaid.org) up to the end of August 2020. Low quality (containing undetermined amino acids, X) or incomplete sequences were excluded. The final number of complete sequences analyzed was 86,163. These were globally aligned using Clustal Omega v1.2.4; the resulting MSA is available at our lab page (https://dbustoslab.github.io/). Then, we calculated the Shannon Entropy to condense the MSA information in a plot (Fig. 1C). The graph showed the second disordered region as highly variable, making it especially attractive to study protein sequence variations, their effects on 14-3-3 binding and viral success. Within this region, the most prominent peak in the entropy plot of the SARS-CoV-2 N protein corresponded to sites 203 and 204, which are among the most variable of the complete SARS-CoV-2 proteome, which can be observed at the GISAID website (GISAID - NextStrain). These sites, together with S194 and S197 were marked with asterisks (Fig. 1C) because they are critical residues that will be discussed below.

The first and second disordered regions contained putative 14-3-3 binding sites (one and three sites respectively). These sites were identified using the SARS-CoV-2 reference amino acidic sequence (EPI_ISL_412977) as input in the ELM resource (http://elm.eu.org/),

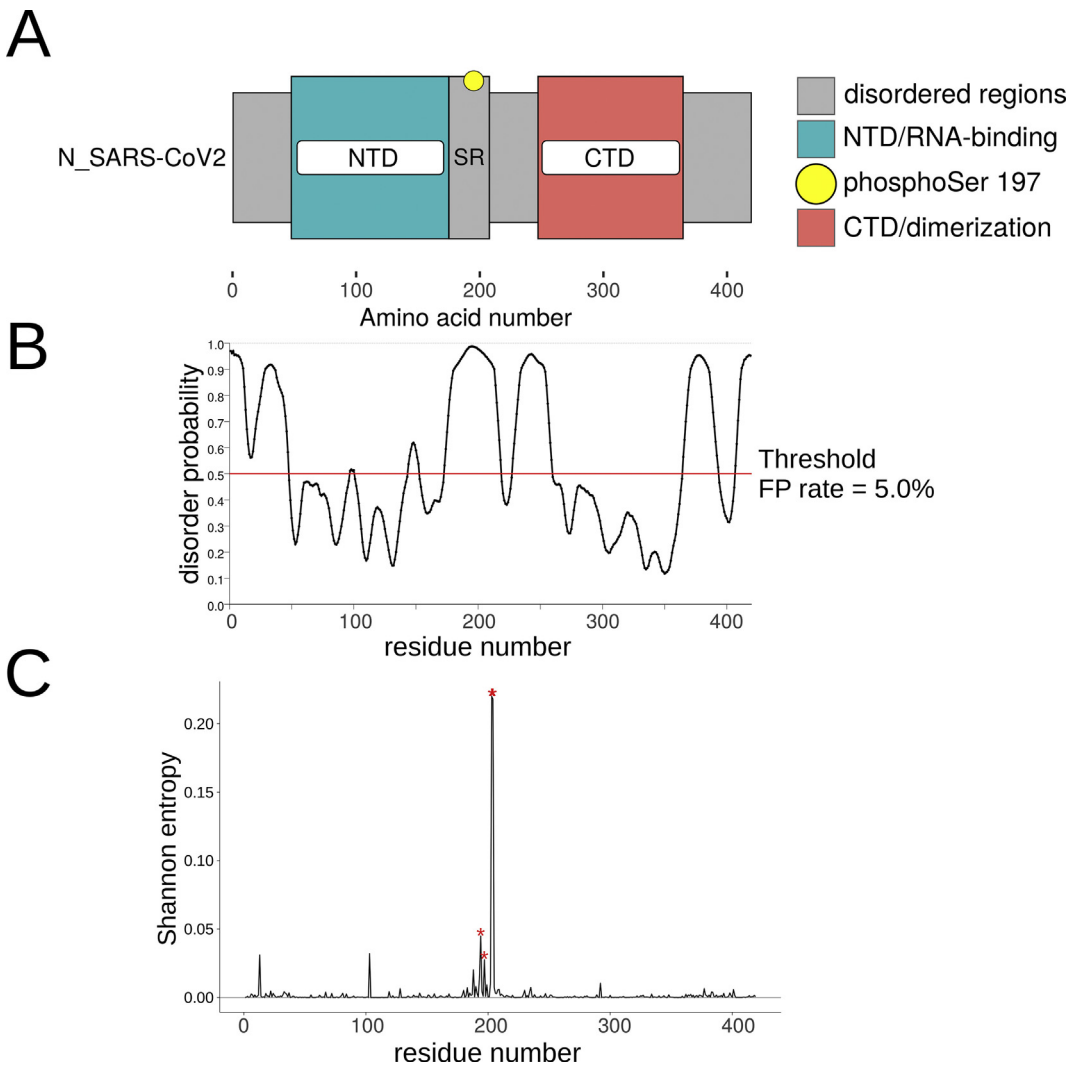


Fig. 1. The SARS-CoV-2 Nucleocapside (N) protein organization, intrinsic disorder and variability. SARS-CoV-2 N protein domains and organization (A), disorder prediction (B), and amino acidic sequence Shannon Entropy prediction (C).

and through visual inspection. Another possible 14-3-3 binding site, covering amino acids 259 to 268, was located inside the structured CTD domain (Fig. 1), and because the binding to 14-3-3 requires structural intrinsic disorder [17], we excluded it from our analysis. Hence, we analyzed sequence variations on the two first disordered regions, specifically in residues contained or nearby the putative 14-3-3 binding sites. These corresponded to amino acids 13 to 19 in the first disordered region and 190 to 204 in the second and were labeled site 1 and 2 respectively (Fig. 2A and B). The latter fragment was located within the SR-rich region previously studied [6,7] and contained three possible concentrically arranged 14-3-3 binding sites of different lengths, the shortest inside the middle-sized and this one inside the longest. This fact, and visual comparison with optimal motifs [1], made us hypothesize that the second site could be the most important in the binding of N protein to 14-3-3 proteins. Indeed, a recent publication [6] demonstrated the critical role of pS197, contained in the SR-rich region (Fig. 1A and B) of the second disordered region, thus we later focused more specifically on mutations of this residue and its neighbors. Remarkably, the S197 showed lower levels of Shannon Entropy as compared to other residues in the SR-rich region (Fig. 1C).

Fig. 2A, B shows amino acid variations in the above-mentioned sites 1 and 2 of SARS-CoV-2 N protein in comparison to the homologous sequence in SARS-CoV1 (GenBank accession number AY274119.3 [3]). Each site including variations was assigned the value 1, which was multiplied by the corresponding score in the BLOSUM62 matrix. We observed that most of the non-conservative mutations were on the red side of the scale, meaning that the changes were naturally unexpected. Although only one site in these two regions was found strictly conserved (F17, Fig. 2A and B), the percentage of strains bearing each mutation is low (less than 3%), except for sites 203 and 204, which showed a significant percentage (36%) of strains bearing both mutations, R203K and G204R (Fig. 2C). Except for these two sites, the analyzed regions were highly conserved. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Fig. 2D and E shows the co-evolution of the SARS-CoV-2 N protein amino acids, computed using two alternative methods (OMES and DCA). The OMES measures the co-variation in the composition of the residues in any two columns in the MSA. The DCA method additionally infers spatial proximity between residues. Both methods identified co-evolving amino acids within the SARS-CoV-2 N protein. However, only OMES found co-evolution involving one of the residues in the known binding motif of N protein to 14-3-3: the S197, critical in this interaction [6], is evolutionarily linked to R203 and G204. The DCA method recognized a strong co-evolution between R203 and G204 but not with

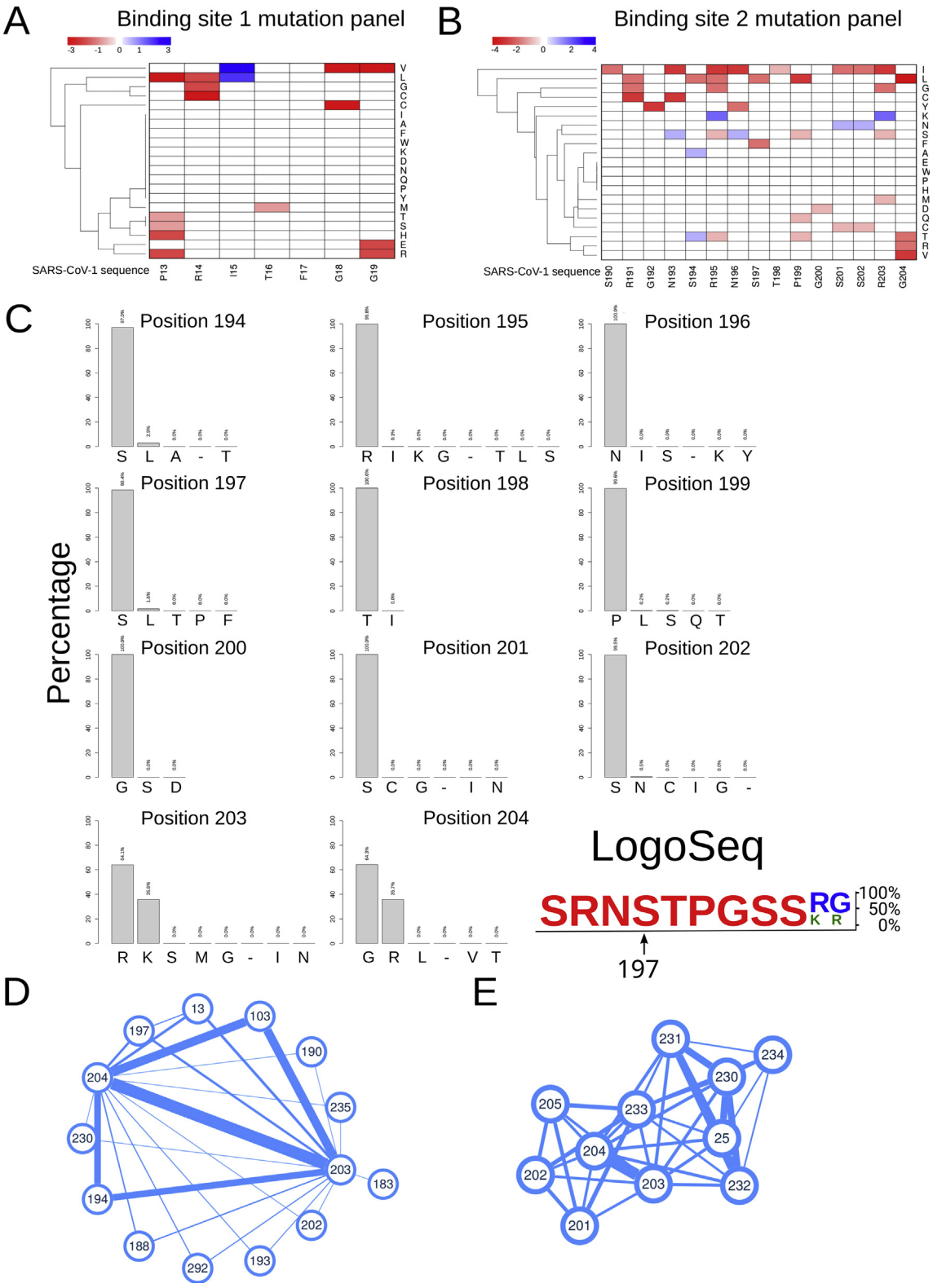


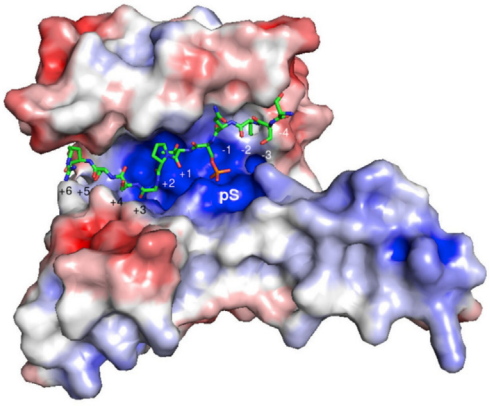
Fig. 2. SARS-CoV-2 N protein variations in disordered regions containing 14-3-3 binding sites and co-evolution analysis. A, B) Heat maps showing amino acid mutations in two selected disordered regions of the SARS-CoV-2 N protein containing putative 14-3-3 binding sites. C) Percentages of amino acid variations in sites surrounding S197, the most important in 14-3-3 binding. D, E) Co-evolution network of N protein amino acids calculated by OMES (D), and DCA method (E). Lines thickness are proportional to the co-evolution score (minimum = 0, maximum = 1).

S197. Based on these methods characteristics, the existence of a functional relationship (but not structural constrain) between residues S197 and R203/G204 is possible.

Although several structures of the N protein NTD and CTD are available (https://www.uniprot.org/uniprot/P59595), there is no data about the SR-rich region including the demonstrated 14-3-3 binding site, probably because of its highly disordered nature. Thus, to study the effect of different mutations on the binding affinity to 14-3-3, we used crystallographic structures of binding peptides with 14-3-3γ, the isoform with the highest binding affinity to the SARS-CoV-2 N protein [6]. Fig. 3 shows this structure and the mscm-ppi2 prediction for the replacement of each position of the 14-3-3 binding peptide by the residues present in that position in SARS-CoV-2 N protein sequences. We observed that those changes positively affecting the predicted interaction between the peptide and 14-3-3 were the ones with the highest prevalence in the phylogenetic tree. In other words, the strains bearing them were the most successfully spread. One of the most interesting examples was observed for mutations in the residue 203, which shows the maximum entropy within the N protein, and even the highest considering the complete SARS-CoV-2 genome (Fig. 4A). The presence of this residue, R203 in the earlier SARS-CoV-2 sequences (Fig. 4B), and also shared with the former SARS-CoV-1, increases the binding to 14-3-3 compared to the T in position +6 in the 14-3-3 binding peptide (Fig. 3). Substitutions to amino acids G, K, M, and S were observed in SARS-CoV-2 N protein amino acid sequences in this site. Among these, only the change to K had a positive effect on the N-protein-14-3-3 interaction, compared to the original T on the 14-3-3 binding peptide. This mutation is the one that has spread most successfully, as it can be observed by its prevalence in the phylogenetic tree (Fig. 4B). The mutation in the contiguous site, G204R, showed a similar pattern on the phylogenetic tree (not shown). Indeed, we found strong evidence of co-evolution between these two sites 203 and 204 (Fig. 2D and E). Using FoldX [15], we observed that either of these two mutations occurring separately (R203K or G204R) destabilized 14-3-3-N protein interaction (R203K, interaction energy 30.08 kcal/mol, and G204R, 30.36 KCal/mol, compared to the wild-type, R203/G204, 35.35 kcal/mol). Intriguingly, the double mutant R203K G204R had compensatory effects, with an interaction energy of 32.22 kcal/mol. This double mutation was rapidly spread after its first appearance (Fig. 4B), and we found that it is present in half of the concern/interest variants, Alpha, Gamma, Zeta, Theta and Lambda. In contrast, the replacement of pS197 by L or F as occurred in a few SARS-CoV-2 variants (Fig. 2) led to impaired binding to 14-3-3 proteins (Fig. 3), and these strains have not been efficiently spread, as it can be observed in Fig. 4C.

Discussion

Our data showed that although the N protein of SARS-CoV-2 is variable along its sequence (Fig. 1C), the region containing the S197 residue, critical for 14-3-3 binding [6] is relatively conserved, except for residues 203 and 204. These two appeared conjointly during December 2019 (https://www.gisaid.org/phylodynamics/global/nextstrain/), which correlates with our result of co-evolution between them (Fig. 2C and D). Interestingly, both mutations seem to drive the phylogenetic tree (amino acids at site 203 were mapped on the tree of Fig. 4B, the same for site 204 was identical, not shown) and match one of the current main clades at the GISAID site, named GR or 20B. For information about how the different clades are defined and named see Ref. [29]. According to previous studies, both mutations drop protein stability [19,20] while favoring torsion [19] or decrease overall structural flexibility



res-in-crystal	res-position	res-in-N-protein	res-number	mcsm-ppi2-prediction	affinity
VAL	-4	SER	193	-0.048	Decreasing
VAL	-4	CYS	193	-0.048	Decreasing
THR	-3	SER	194	0.06	Increasing
THR	-3	LEU	194	-0.218	Decreasing
SER	-2	THR	195	-0.445	Decreasing
SER	-2	ARG	195	0.123	Increasing
SER	-1	TYR	196	0.179	Increasing
SER	-1	LYS	196	0.071	Increasing
SER	-1	ASN	196	-0.173	Decreasing
pSER	0	LEU	197	---------	No binding
pSER	0	PHE	197	---------	No binding
CYS	1	THR	198	0.05	Increasing
PRO	2	GLN	199	0.123	Increasing
PRO	2	LEU	199	-0.032	Decreasing
PRO	2	SER	199	-0.078	Decreasing
ALA	3	ASP	200	0.649	Increasing
ALA	3	GLY	200	-0.03	Decreasing
ASP	4	SER	201	-0.076	Decreasing
LEU	5	SER	202	-0.848	Decreasing
THR	6	ARG	203	0.045	Increasing
THR	6	GLY	203	-0.162	Decreasing
THR	6	LYS	203	0.108	Increasing
THR	6	MET	203	-0.041	Decreasing
THR	6	SER	203	-0.123	Decreasing

Fig. 3. Mutations on the SR-rich region of N protein affect the binding potential to 14-3-3 of SARS-CoV-2 variants. Effect of S197 and surrounding amino acids mutations in N protein on the binding potential to host 14-3-3γ. The crystal Structure (PDBID: 2B05) of 14-3-3γ in complex with a phosphoserine peptide was single mutated to each amino acid of SARS-CoV-2 N protein in the corresponding position. The binding energy was calculated for each change, and determined if it was stabilizing or destabilizing by comparison with the original peptide in the crystal structure.

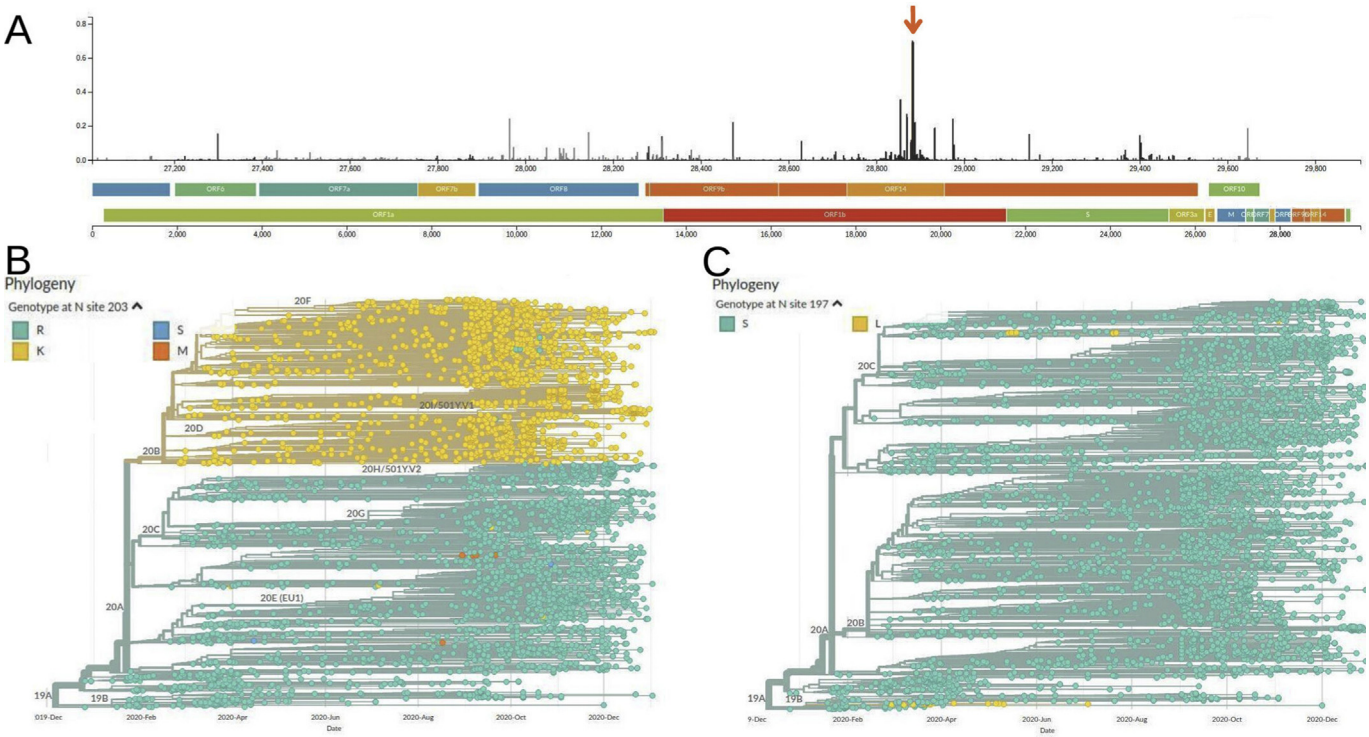


Fig. 4. General variability of the SARS-CoV-2 genome and phylogenetic distribution of sites 203 and 197. A) Bars graph (GISAID, https://www.gisaid.org) showing the entropy of the different residues along the SARS-CoV-2 genome. The bars corresponding to residues 203 and 204 were indicated with an arrow. B) Phylogeny (GISAID) showing mutations of site 203 to different amino acids with different levels of dissemination within the tree. C) Same as in B except that site 197 phylogenetic distribution is represented.

[20]. More recently, Vahed et al. [21] found that the mutant type (K203, R204) showed higher free energy at these residues, suggesting enhanced structural flexibility. Besides these observations, we found that any of these two mutations occurring separately destabilized the interaction between 14-3-3 and the N protein, while the double mutant R203K G204R showed compensatory effects. This linked functionally these residues to the S197, the most important residue for the binding to 14-3-3 proteins [6]. In the 14-3-3-N protein interaction, the 203 residue accommodates near the border of the 14-3-3 binding grove (position +6, Fig. 3). Although 204 residue stays outside the 14-3-3 binding grove, we have previously identified essential residues for 14-3-3 interactions with this feature [22]. The SARS-CoV-2 N protein R203K/G204R double mutant has successfully spread, whereas single R203K or G204R mutants did not spread at all. As explained earlier, most strains bearing the R203K mutation also include the G204R mutation.

The current SARS-CoV-2 variants that raise the risk to public health have been recently classified (Tracking SARS-CoV-2 variants (who. int)) as either of concern (VOC): Alpha (B1.1.7, first documented in the UK, Sept. 2020), Beta (B1.351, SA, May 2020), Gamma (P1, Brazil, Nov. 2020), and Delta (B1.617.2, India, Oct. 2020), or of interest (VOI): Epsilon (B1.427/9, USA, Mar. 2020), Zeta (P2, Brazil, Apr. 2020), Eta (B1.525, multiple countries, Dec. 2020), Theta (P3, Philippines, Jan. 2021), Iota (B1.526, USA, Nov. 2020), Kappa (B1.617.1, India, Oct. 2020), and Lambda (C37, Peru, Aug. 2020). Thus, we compared the data presented with the consensus sequences of these circulating variants. We found that half of them (Alpha, Gamma, Zeta, Theta, and Lambda) include the mutation R203K, always linked to the G204R mutation. Only two of them (Delta and Kappa) contain in this site a different mutation, R203 M, however, this modification is not linked to any variation in position 204, which remains as glycine. This may be related to the fact that the change to K in position +6 in the crystal (which corresponds to residue 203 in N protein) increases affinity to 14-3-3 proteins, with a positive mscm-ppi2-prediction value of 0.108 (Fig. 3). In contrast, the change to M is slightly negative (Figs. 3, -0.041), thus it decreases the binding affinity. As expected, the VOCs and VOIs do not show other mutations in the 14-3-3 binding groove, from residues 193 to 203. This agrees with our results, and again supports the idea that this region of the N protein is highly conserved, and that the interaction N protein-14-3-3 is an appealing target for the development of antiviral tools.

Besides, to investigate the main groove amino acid variations of the 14-3-3 protein family in different human ethnic groups we used the information from the gnomAD database [23]. Any possible loss-of-function analysis requires large sample sizes. The gnomAD database contains 125,748 exomes and 15,708 genomes from healthy humans of different ethnics backgrounds around the world. No 14-3-3 isoforms, including gamma isoform, which is the one with the strongest interaction with the N protein [6], accumulate mutations in the important residues for the binding to their partners.

Many clues suggest the biological importance of 14-3-3-N protein interaction. First, 14-3-3 isoforms bind the N protein with different affinities, with 14-3-3γ being the strongest [8]. We have previously studied the interaction networks of all human 14-3-3 isoforms and found that the 14-3-3γ interactome is specifically enriched in RNA processing proteins [11]. This isoform has been reported in P-bodies, structures specialized in translational repression, mRNA surveillance, and degradation. Its depletion by siRNA showed 14-3-3γ as a regulator of P-body formation [24]. The N protein is well known to form liquid condensates through binding specific sequences from its genome [25]. Liquid-Liquid Phase Separation (LLPS) induced by viral proteins are considered

"viral factories", where viral replication and assembly take place [26]. We propose that the SARS-CoV-2 uses 14-3-3 as part of its life cycle, for replication and/or assembly, through the interaction of N with 14-3-3γ proteins within liquid condensates. Another possibility is that LLPSs are not associated with viral replication and assembly, and rather interfere with host cell functions through alteration of gene transcription or interaction with cellular proteins [26]. If this is the case for SARS-CoV-2, the N protein could be "sequestering" 14-3-3s and disrupting their numerous functions, as previously suggested [6]. Among other examples, 14-3-3s are involved in autophagy [1], and other host cell defense mechanisms [6]. A third hypothesis is that 14-3-3 binding may regulate N protein shuttling, as seen previously for SARS-CoV-1 N protein [4]. In favor of this are the reported enhancement of nuclear localization and export Signals in Coronavirus associated with high case fatality reports, including SARS-CoV-2 [27].

Phosphorylation regulates the state and function of the N protein. The unmodified protein would be in a gel state for the nucleocapsid, while the phosphorylated version would be in a liquid state to process the viral genome [28]. This is in agreement with our results, which suggest that binding of N protein to 14-3-3 is beneficial for the virus and supports the hypothesis that either the virus uses 14-3-3 proteins to complete its life cycle, or it arrests them to impede their normal functions in the host cell. In any case, targeting this N protein-14-3-3 viral-host protein-protein interaction seems an attractive approach for exploring antiviral strategies. Future studies should be oriented in that direction.

Author contributions

DMB and MU conceived and supervised the study, analyzed data and wrote the manuscript; SDV, LR, DMB, and MU designed and performed experiments; SDV and LR contributed equally.

Role of the funding source

The funding institution had no role in study design; collection, analysis, and interpretation of data; in writing the report; and in the decision to publish.

Declaration of competing interest

The authors declare that there is no conflict of interest.

Acknowledgments

SDV and LR hold Ph.D. scholarships from CONICET, Argentina; DMB and MU are CONICET Researchers. This work was supported by ANPCYT, Argentina (grant PICT'17–1984). We thank all the researches that submitted the sequences used.

References

[1] N.N. Sluchanko, D.M. Bustos, Intrinsic disorder associated with 14-3-3 proteins and their partners, in: Danc. Protein Clouds Intrinsically Disord. Proteins Norm Pathol., First Ed., Elsevier Inc., 2019, pp. 19–61.
[2] K.G. Nathan, S.K. Lal, The multifarious role of 14-3-3 family of proteins in viral replication, Viruses 12 (2020) 436, https://doi.org/10.3390/v12040436.
[3] M.A. Marra, S.J.M. Jones, C.R. Astell, et al., The Genome Sequence of the SARS-Associated Coronavirus, Science 300 (2003) 1399–1405, https://doi.org/10.1126/science.1085953.
[4] M. Surjit, R. Kumar, R.N. Mishra, et al., The Severe Acute respiratory Syndrome Coronavirus nucleocapsid protein is phosphorylated and localizes in the cytoplasm by 14-3-3-mediated translocation, J. Virol. 79 (2005) 11476–11486, https://doi.org/10.1128/jvi.79.17.11476-11486.2005.
[5] Y. Cong, M. Ulasli, H. Schepers, et al., Nucleocapsid protein recruitment to replication-transcription complexes plays a crucial role in coronaviral life cycle, J. Virol. 94 (2019), https://doi.org/10.1128/jvi.01925-19 e01925–19.
[6] K.V. Tugaeva, D.E.D.P. Hawkins, J.L. Smith, et al., The mechanism of SARS-CoV-2 nucleocapsid recognition by the human 14-3-3 proteins, J. Mol. Biol. 433 (2021) 166, https://doi.org/10.1016/j.jmb.2021.166875.
[7] H.Y.L. Tung, P. Limtung, Mutations in the phosphorylation sites of SARS-CoV-2 encoded nucleocapsid protein and structure model of sequestration by protein 14-3-3, Biochem. Biophys. Res. Commun. 532 (2020) 134–138, https://doi.org/10.1016/j.bbrc.2020.08.024.
[8] S. Elbe, G. Buckland-Merrett, Data, disease and diplomacy: GISAID's innovative contribution to global health, Glob. Chall. 1 (2017) 33–46, https://doi.org/10.1002/gch2.1018.
[9] O. Bastien, S. Roy, É. Maréchal, Construction of non-symmetric substitution matrices derived from proteomes with biased amino acid distributions, Comptes Rendus Biol. 328 (2005) 445–453, https://doi.org/10.1016/j.crvi.2005.02.002.
[10] I. Walsh, A.J.M. Martin, T. Di domenico, et al., Espritz: accurate and fast prediction of protein disorder, Bioinformatics 28 (2012) 503–509, https://doi.org/10.1093/bioinformatics/btr682.
[11] M. Uhart, D.M. Bustos, Human 14-3-3 paralogs differences uncovered by cross-talk of phosphorylation and lysine acetylation, PloS One 8 (2013) e55703, https://doi.org/10.1371/journal.pone.0055703.
[12] J. Pelé, B. Taddese, M. Deniaud, et al., Co-variation approaches to the evolution of protein families, Adv. Tech. Biol. Med. 5 (2017) 250–254, https://doi.org/10.4172/2379-1764.1000250.
[13] P.M. Valero-Mora, ggplot2: elegant graphics for data analysis, J. Stat. Software 35 (2010) 1–3, https://doi.org/10.18637/jss.v035.b01.
[14] D.E.V. Pires, D.B. Ascher, T.L. Blundell, MCSM: predicting the effects of mutations in proteins using graph-based signatures, Bioinformatics 30 (2014) 335–342, https://doi.org/10.1093/bioinformatics/btt691.
[15] J. Delgado, L.G. Radusky, D. Cianferoni, et al., FoldX 5.0: working with RNA, small molecules and a new graphical interface, Bioinformatics 35 (2019) 4168–4169, https://doi.org/10.1093/bioinformatics/btz184.
[16] M.B. Zerihun, F. Pucci, E.K. Peter, et al., Pydca v1. 0: a comprehensive software for Direct coupling analysis of RNA and protein sequences, Bioinformatics 36 (2020) 2264–2265, https://doi.org/10.1093/bioinformatics/btz892.
[17] D.M. Bustos, A.A. Iglesias, Intrinsic disorder is a key characteristic in partners that bind 14-3-3 proteins, Proteins Struct. Funct. Genet. 63 (2006) 35–42, https://doi.org/10.1002/prot.20888.
[19] D. Benvenuto, A.B. Demir, M. Giovanetti, et al., Evidence for mutations in SARS-CoV-2 Italian isolates potentially affecting virus transmission, J. Med. Virol. 92 (2020) 2232–2237, https://doi.org/10.1002/jmv.26104.
[20] M.S. Rahman, M.R. Islam, A.S.M.R.U. Alam, et al., Evolutionary dynamics of SARS-CoV-2 nucleocapsid protein and its consequences, J. Med. Virol. 93 (2020) 2177–2195, https://doi.org/10.1002/jmv.26626.
[21] M. Vahed, T.M. Calcagno, E. Quinonez, et al., Impacts of 203/204: RG>KR mutation in the N protein of SARS-CoV-2 1, BioRxiv (2021), https://doi.org/10.1101/2021.01.14.426726.
[22] M. Uhart, A.A. Iglesias, D.M. Bustos, Structurally constrained residues outside the binding motif are essential in the interaction of 14-3-3 and phosphorylated partner, J. Mol. Biol. 406 (2011) 552–557, https://doi.org/10.1016/j.jmb.2010.12.043.
[23] K.J. Karczewski, L.C. Francioli, G. Tiao, et al., The mutational constraint spectrum quantified from variation in 141,456 humans, Nature 581 (2020) 434–443, https://doi.org/10.1038/s41586-020-2308-7.
[24] N. Okada, N. Yabuta, H. Suzuki, et al., A novel Chk1/2–Lats2–14-3-3 signaling pathway regulates P-body formation in response to UV damage, J. Cell Sci. 124 (2010) 57–67, https://doi.org/10.1242/jcs.072918.
[25] C. Iserman, C.A. Roden, M.A. Boerneke, et al., Genomic RNA elements drive phase separation of the SARS-CoV-2 nucleocapsid, Mol. Cell. 80 (2020) 1078–1091, https://doi.org/10.1016/j.molcel.2020.11.041.
[26] S. Brocca, R. Grandori, S. Longhi, et al., Liquid–liquid phase separation by intrinsically disordered protein regions of viruses: roles in viral life cycle and control of virus–host interactions, Int. J. Mol. Sci. 21 (2020), 9045, https://doi.org/10.3390/ijms21239045.
[27] A.B. Gussow, N. Auslander, G. Faure, et al., Genomic determinants of pathogenicity in SARS-CoV-2 and other human coronaviruses, Proc. Natl. Acad. Sci. U.S.A. 117 (2020) 15193–15199, https://doi.org/10.1073/pnas.2008176117.
[28] C.R. Carlson, J.B. Asfaha, C.M. Ghent, et al., Phosphoregulation of phase separation by the SARS-CoV-2 N protein suggests a biophysical basis for its dual functions, Mol. Cell. 80 (2020) 1092–1103, https://doi.org/10.1016/j.molcel.2020.11.025, e4.
[29] https://nextstrain.org/blog/2020-06-02-SARSCoV2-clade-naming, 2020.