SARS-CoV-2 Nsp3 Unique Domain SUD interacts with Guanine quadruplexes and G4-ligands inhibit this interaction.

M. Lavigne[1]*, O. Helynck[1], P. Rigolet[2], R. Boudria-Souilah[1], M. Nowakowski[3], B. Baron[4], S. Brûlé[4], S. Hoos[4], B. Raynal[4], L. Guittat[5], C. Beauvineau[2], S. Petres[3], A. Granzhan[2], J. Guillon[6], G. Pratviel[7], M-P. Teulade-Fichou[2], P. England[4]*, J-L Mergny[8]*, H. Munier-Lehmann[1]*

SUPPLEMENTARY DATA

Supplementary Table 1. Potential G4 forming sequences (G4FS) of the SARS-CoV-2 genome identified by Panera *et al*. (1) and Ji *et al*. (2), according to QGRS mapper (3); the corresponding G-scores provided by the authors are shown. For comparison, we calculated the G4-Hunter scores according to Bedrat *et al*. (4) in the right most column. These scores are low (all below 1.05 and sometimes below 0.5, for which G4 formation has never been reported before) and indicative of unlikely or unstable G4 formation. Scores above 1.0 are indicated in bold characters. For comparison, the G4-Hunter score of the thrombin binding aptamer (TBA), one of the shortest intramolecular G4 (15nt) with two stacked quartets, is 1.13. Three of these G4FRG1, RG2, RG3 (Zhao et al., 2020) have been used in our study.

Supplementary Table II. Oligonucleotide sequences used to clone SARS-CoV-2 SUD domains in pET28a+ and to mutagenize its sequence (MutA).

Supplementary Figure S1. Primary sequence of purified SARS-CoV-2 proteins. Purified proteins share a polyhistidine tag and a thrombin proteolytic cleavage site at their N-terminal part (colored in red).

Supplementary Figure S2. Quality control results of SARS-CoV-2 SUD-M. a) represents the mono and multi-charged ions of SUD-M acquired in MALDI-TOF. SUD-M loses its first methionine; **b)** shows a typical UV spectrum of a protein without contaminations and scattering at 260 nm and 340 nm, respectively; **c)** and **d)** shows SUD-M in the best buffer condition determined by nanoDSF and DLS, respectively. Both techniques highlight that SUD-M is best behaved in presence of 20 mM Tris-HCl pH 9, 500 mM NaCl.

Supplementary Figure S3. Quality control results of SARS-CoV-2 SUD-NM. a) represents the multi-charged ions of SUD-NM acquired in MALDI-TOF. SUD-NM loses its first methionine; **b)** shows a typical UV spectrum of a protein without contaminations and scattering at 260 nm and 340 nm, respectively; **c)** and **d)** shows SUD-NM in the best buffer condition determined by nanoDSF and DLS, respectively. Both techniques highlight that SUD-NM is best behaved in presence of 20 mM Tris-HCl pH 9, 500 mM NaCl.

Supplementary Figure S4. Quality control results of SARS-CoV-2 SUD-NMC. **a)** represents the mono and multi-charged ions of SUD-NMC acquired in MALDI-TOF. SUD-NMC loses its first methionine; **b)** shows a typical UV spectrum of a protein without contaminations and scattering at 260 nm and 340 nm, respectively; **c)** and **d)** shows SUD-NMC in the best buffer condition determined by nanoDSF and DLS, respectively. Both techniques highlight that SUD-NMC is best behaved in presence of 20 mM Tris-HCl pH 9, 500 mM NaCl.

Supplementary Figure S5. Circular Dichroism (CD) and Isothermal Differential Spectra (IDS) of RNA TRF2 (panels A & B) and DNA BCL2 (panels C & D) oligonucleotides.

CD spectra are shown on the left (panels A & C) while IDS are shown on the right (panels B & D). “Bio” stands for biotin. WT sequences are the original G4-prone motifs, while “Mut” indicates the oligonucleotides in which base substitutions have been introduced to hamper G4 formation.

Supplementary Figure S6. SARS-CoV-2 SUD-NM interaction with the c-myc G4s (a-b) and effect of the Trimethyl-Psoralen on the SUD-NM / TRF2 G4 interaction reported by the HTRF assay a) His-SUD-NM (30 nM) and biotinylated c-MYC (0-500 nM) were incubated for 1h at room temperature before addition of the donor and acceptor conjugates **b)** His-SUD-NM (30 nM) and biotinylated c-MYC (200 or 400 nM) were incubated for 1h at room temperature in the presence of increasing concentrations of non-biotinylated c-MYC WT (0 to 10 µM). **c)** His-SUD-NM (10 nM) and biotinylated TRF2 G4 (50 nM) were incubated for 1 h at room temperature in the presence of increasing concentrations (from 100 to 500 nM) of TMP.

Supplementary Figure S7. Docking of the human telomeric RNA (TERRA) quadruplex (PDB code: 3IBK (5)) on the modeled structure of SARS-CoV-2 SUD-NM. The modeled structure of SARS-CoV-2 SUD-NM is presented as an electrostatic surface, whereas the G4-RNA structure is figured in orange and in sticks representation with the potassium ions depicted as purple spheres. **a)** Electrostatic surface of the modeled SUD monomer showing positively charged regions (in blue) on the N and M macrodomains **b)** The best score pose of the 3IBK G4-RNA structure docked on the monomer of the modeled structure of SARS-CoV-2 SUD-NM. The putative binding site is a depression between the SUD-N and SUD-M macrodomains, which is partially positively charged. c) Same as b) but turned 90° along the vertical axis. d) Electrostatic surface of the modeled SUD-NM dimer showing an extended positively charged region running on the dimer surface. e) Docking of the 3IBK G4-RNA structure on the dimer of the modeled structure of SARS-CoV-2 SUD-NM. f) Same as e) but turned 90° along the vertical axis.

Supplementary Figure S8: a-b) Regions and secondary structure elements (SSEs) shaping the putative binding sites of G4 structures in SUD-NM and dimer interface. a) G4-DNA (PDB code 2F8U) in the dimer of the modeled SUD-NM from SARS-CoV-2. The SUD-NM protein is in cartoon representation and the G4 structure is figured in orange and in sticks representation. SUD-N appears in blue and SUD-M in red. The SSEs of SUD-N involved in the putative binding sites are colored in cyan and the SSEs of SUD-M involved in the putative binding sites are in magenta. The residues that have been mutated in Mut4 are labeled in stick representation with carbons in yellow. **(b)** Residues involved in the dimer interface of the modeled structure of SUD-NM from SARS-CoV-2. The residues mutated in the dimer interface (MutA) are in stick representation with carbons in yellow and residue E643, in interaction, appears with carbons in green. **c-d) HTRF assay measuring the interaction between the WT, MutA or Mut4 SARS-CoV-2 SUD-NM proteins and the BCL2 or TRF2 G4s**. WT or Mut4 His-SUD-NM protein (3-10-30 nM) were incubated with 50 nM of biotinylated BCL2 WT or TRF2 WT for 1h at room temperature before addition of the donor and acceptor conjugates. HTRF ratios were measured as described in Material and Methods.

Supplementary Figure S9: Structural model of the SARS-CoV-2 SUD-NM dimer complexed with Paip1 and a G4-RNA. The modeled SARS-Cov2 SUD-NM (from this study) and the crystallographic human Paip1 proteins (6) are in cartoon representation and the G4-RNA structure is figured in orange and in sticks representation. SUD-N appears in blue, SUD-M in red, the linker between SUD-N and SUD-M in grey and the Paip1 protein in green. RMSD=0.62 A for the superposition of the C-alpha atoms of SUD-N in SARS-Cov-SUD-N-Paip1 complex (6) and SUD-N in SUD-NM from SARS-Cov2 (this study). RMSD=0.25 A for 940 atoms aligned in the superposition of SUD-N from the crystallographic structures of the SARS-CoV SUD-NM (pdb code: 2w2g) and the SARS-Cov-SUD-NM-Paip1 complex (pdb code: 6yxj).

SUPPLEMENTARY METHODS

Method for quality control of the recombinant SUD proteins

The method is adapted from a previously published approach (7) and ARBRE-MOBIEU P4EU recommendations (https://arbre-mobieu.eu/guidelines-on-protein-quality-control/).

Assessment of the sample homogeneity and quantification:

Briefly, DLS analysis was performed on a DynaPro Plate Reader III (Wyatt, Santa Barbara, CA, USA) to ensure that the samples did not contain aggregates. Experiments were performed in triplicate in a 384-well microplate (Corning ref 3540, New-York, USA), with 20 acquisitions of 10 s each at 20°C, monitored with the DYNAMICS version V7.10.0.21 software (Wyatt, Santa Barbara, CA, USA).

Detection of DNA contamination at 260 nm and scattering at 340 nm. Protein quantification at 280 nm was carried out by recording a full spectrum between 240 and 340 nm. Measurements were done at 20°C in a 1 cm quartz cell, reference 105.202-QS.10 (Hellma, France), using a JASCO V-650 spectrophotometer (JASCO Corporation, Japan). A baseline subtraction at 340 nm was performed with the Spekwin32 software (F. Menges "Spekwin32 - optical spectroscopy software", Version 1.72.2, 2016, http://www.effemm2.de/spekwin/) to accurately calculate the protein concentration.

Assessment of the integrity of samples

Protein integrity and purity were analyzed on a Bruker UltrafeXtreme MALDI-TOF/TOF instrument. A volume of 1µL of protein at 0.3mg/ml was eluted on a MTP 384 ground steel target plate (Bruker-Daltonics, Germany) with 1 µl of a 20 mg/ml α-Cyano-4-hydroxycinnamic acid (HCCA) in 50% acetonitrile (ACN), 0.1 % trifluoroacetic acid (TFA) as matrix solution. Data were acquired using Flexcontrol software (Bruker-Daltonics, Germany) and shots were recorded in positive ion linear mode. Mass spectra were externally calibrated in the *m/z* range of 5-55 kDa with both Protein Standard I and II (Bruker-Daltonics, Germany) and analyzed with the Flexanalysis software (Bruker). Finally, buffer optimization was performed by monitoring the time stability and the thermal stability in 20 buffer conditions by DLS and nanoDSF, respectively.

Buffer optimization

- By DLS: 2 µL of protein were mixed with 20 µL of buffer. Experiments were performed at 20°C for 24 hours with measurement of 20 acquisitions of 10s each. A 1 hour wait was added at the end of each cycle. The DYNAMICS version V7.10.0.21 software (Wyatt, Santa Barbara, CA, USA) was used to monitor and to analyze the data.
- By nanoDSF: thermal denaturation was performed on the Prometheus NT.48 instrument (Nanotemper) from 20 to 95°C with a 2°C / min heating rate. The tryptophan fluorescence emission was monitored at 330 nm and 350 nm as a function of increasing temperature. Briefly, 1 µL of protein was diluted in 10 µL of each buffer. The capillaries were filled with the sample and the intrinsic fluorescence signal expressed as the 350 nm/330 nm emission ratio was plotted as a function of temperature.

References

1. Panera, N., Tozzi, A.E. and Alisi, A. (2020) The G-Quadruplex/Helicase World as a Potential Antiviral Approach Against COVID-19. *Drugs*.
2. Ji, D., Juhas, M., Tsang, C.M., Kwok, C.K., Li, Y. and Zhang, Y. (2020) Discovery of G-quadruplex-forming sequences in SARS-CoV-2. *Brief Bioinform*.
3. Kikin, O., D'Antonio, L. and Bagga, P.S. (2006) QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. *Nucleic acids research*, **34**, W676-682.

4. Bedrat, A., Lacroix, L. and Mergny, J.L. (2016) Re-evaluation of G-quadruplex propensity with G4Hunter. *Nucleic acids research*, **44**, 1746-1759.
5. Collie, G.W., Haider, S.M., Neidle, S. and Parkinson, G.N. (2010) A crystallographic and modelling study of a human telomeric RNA (TERRA) quadruplex. *Nucleic Acids Res*, **38**, 5569-5580.
6. Lei, J., Ma-Lauer, Y., Han, Y., Thoms, M., Buschauer, R., Jores, J., Thiel, V., Beckmann, R., Deng, W., Leonhardt, H. *et al.* (2021) The SARS-unique domain (SUD) of SARS-CoV and SARS-CoV-2 interacts with human Paip1 to enhance viral RNA translation. *EMBO J*, e102277.
7. Raynal, B., Lenormand, P., Baron, B., Hoos, S. and England, P. (2014) Quality assessment and optimization of purified protein samples: why and how? *Microb Cell Fact*, **13**, 180.

Length	Sequence	G-score	G4-Hunter	
20	GGTATGTGGAAAGGTTATGG	19	0.85	RG2
24	GGCTTATAGGTTTAATGGTATTGG	19	0.63	RG3
26	GGTGTTGTTGGAGAAGGTTCCGAAGG	18	0.62	
15	GGCTGGCAATGGCGG	18	0.87	RG1
20	GGTTGGACCTTTGGTGCAGG	17	0.60	
22	GGCCATGGTACATTTGGCTAGG	17	0.46	
25	GGCTTTGGAGACTCCGTGGAGGAGG	16	0.64	
20	GGTAATAAAGGAGCTGGTGG	15	0.80	
17	GGAGGAGGTGTTGCAGG	15	**1.00**	
23	GGATACAAGGCTATTGATGGTGG	14	0.65	
19	GGAAATTTTGGGGACCAGG	14	**1.05**	
30	GGCATGGAAGTCACACCTTCGGGAACGTGG	11	0.47	
29	GGCGGTGCACCAACAAAGGTTACTTTTGG	10	0.31	
29	GGGTTTAAATGGTTACACTGTAGAGGAGG	10	0.72	
29	GGTGATTCTTCTTCAGGTTGGACAGCTGG	10	0.45	
30	GGCTGGTAATGTTCAACTCAGGGTTATTGG	9	0.60	
23	GGATTGGCTTCGATGTCGAGGGG	9	**1.04**	
29	GGCTTCTACGCAGAAGGGAGCAGAGGCGG	9	0.66	
29	GGTGGTTATACTGAAAAATGGGAATCTGG	8	0.69	
30	GGATCACCGGTGGAATTGCTATCGCAATGG	7	0.33	
26	GGTGGTCGCACTATTGCCTTTGGAGG	6	0.42	
27	GGTTATACCTACTAAAAAGGCTGGTGG	6	0.37	
27	GGAACAAGCAAATTCTATGGTGGTTGG	6	0.52	
28	GGTTTTCCATTTAATAAATGGGGTAAGG	4	0.71	
29	GGCGGTTCACTATATGTTAAACCAGGTGG	3	0.35	

Supplementary Table I

Name	Sequence (5’ to 3’)
CoV2SUDNForNde	GGCCGCATATGAAGATCAAGGCCTGCGTGGAG
CoV2SUDMForNde	GCCGCATATGATCCTGGGCACCGTGAGCTG
CoV2SUDNRevXho	GCCGCTCGAGTCATTCCTGCTTCTCGTTAGAGATG
CoV2SUDMRevXho	GCCGCTCGAGTCAGCTAGAGGAGGTCAGGTATCC
CoV2SUDCRevXho	GCCGCTCGAGTCAGCGCAGAGACAGCAGGGTC
CoV2SUDmutAFor	ACCAAGAAGGCCGCTGCCGCCACCGAGATGCTGGCCGCGGCTCTGCGTAAG
CoV2SUDmutARev	CTTACGCAGAGCCGCGGCCAGCATCTCGGTGGCGGCAGCGGCCTTCTTGGT

Supplementary Table II

>SUD-M SARS-CoV2, MW: 16464.82 Da, pI: 9.04

MGSSHHHHHHSSGLVPRGSHMILGTVSWNLREMLAHAEETRKLMPVCVETKAIV
STIQRKYKGIKIQEGVVDYGARFYFYTSKTTVASLINTLNDLNETLVTMPLGYV
THGLNLEEAARYMRSLKVPATVSVSSPDAVTAYNGYLTSSS

>SUD-NM SARS-CoV2, MW: 31356.02 Da, pI: 7.15

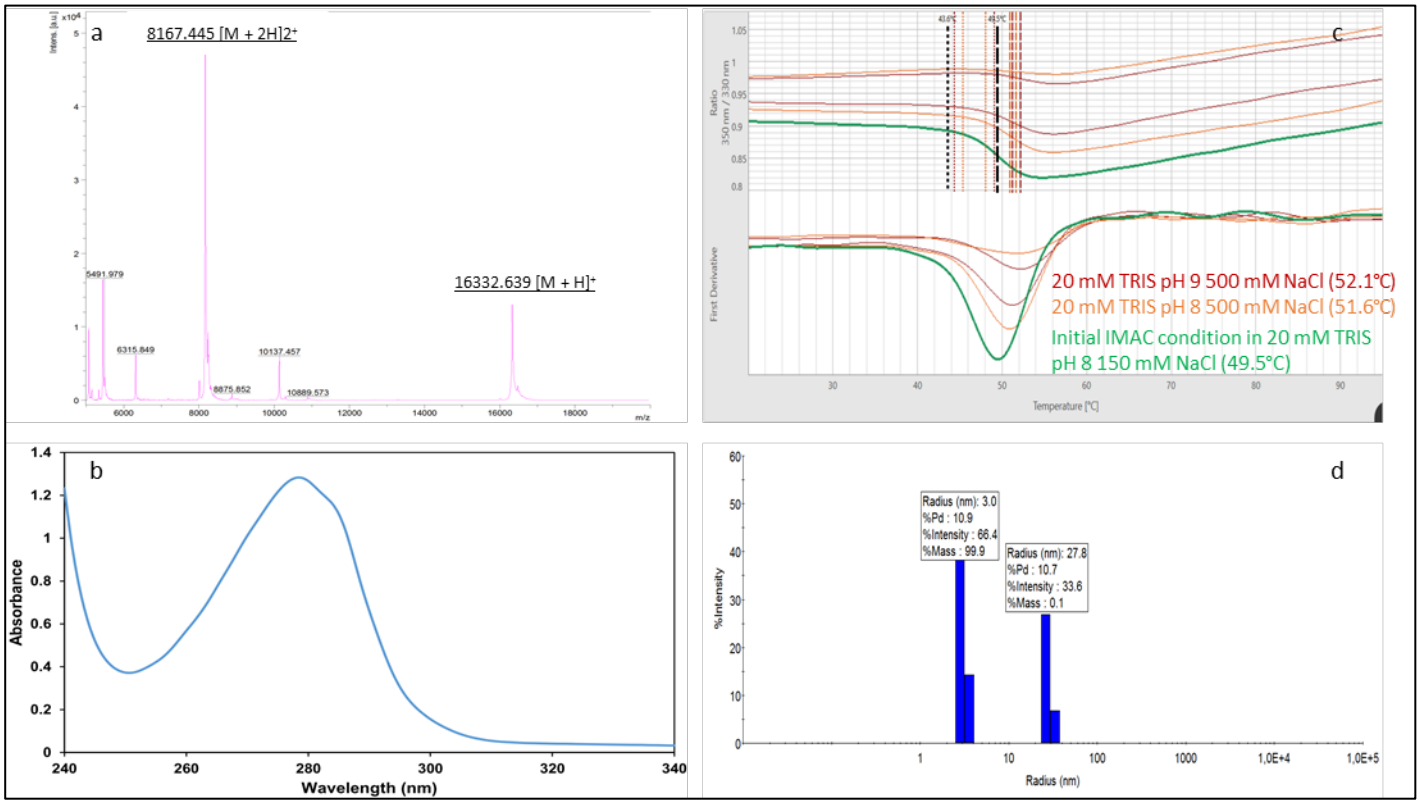
MGSSHHHHHHSSGLVPRGSHMKIKACVEEVTTTLEETKFLTENLLLYIDINGNL
HPDSATLVSDIDITFLKKDAPYIVGDVVQEGVLTAVVIPTKKAGGTTEMLAKAL
RKVPTDNYITTYPGQGLNGYTVEEAKTVLKKCKSAFYILPSIISNEKQEILGTV
SWNLREMLAHAEETRKLMPVCVETKAIVSTIQRKYKGIKIQEGVVDYGARFYFY
TSKTTVASLINTLNDLNETLVTMPLGYVTHGLNLEEAARYMRSLKVPATVSVSS
PDAVTAYNGYLTSSS

>SUD-NMC SARS-CoV2, MW: 39213.79 Da, pI: 6.64

MGSSHHHHHHSSGLVPRGSHMKIKACVEEVTTTLEETKFLTENLLLYIDINGNL
HPDSATLVSDIDITFLKKDAPYIVGDVVQEGVLTAVVIPTKKAGGTTEMLAKAL
RKVPTDNYITTYPGQGLNGYTVEEAKTVLKKCKSAFYILPSIISNEKQEILGTV
SWNLREMLAHAEETRKLMPVCVETKAIVSTIQRKYKGIKIQEGVVDYGARFYFY
TSKTTVASLINTLNDLNETLVTMPLGYVTHGLNLEEAARYMRSLKVPATVSVSS
PDAVTAYNGYLTSSSKTPEEHFIETISLAGSYKDWSYSGQSTQLGIEFLKRGDK
SVYYTSNPTTFHLDGEVITFDNLKTLLSLR

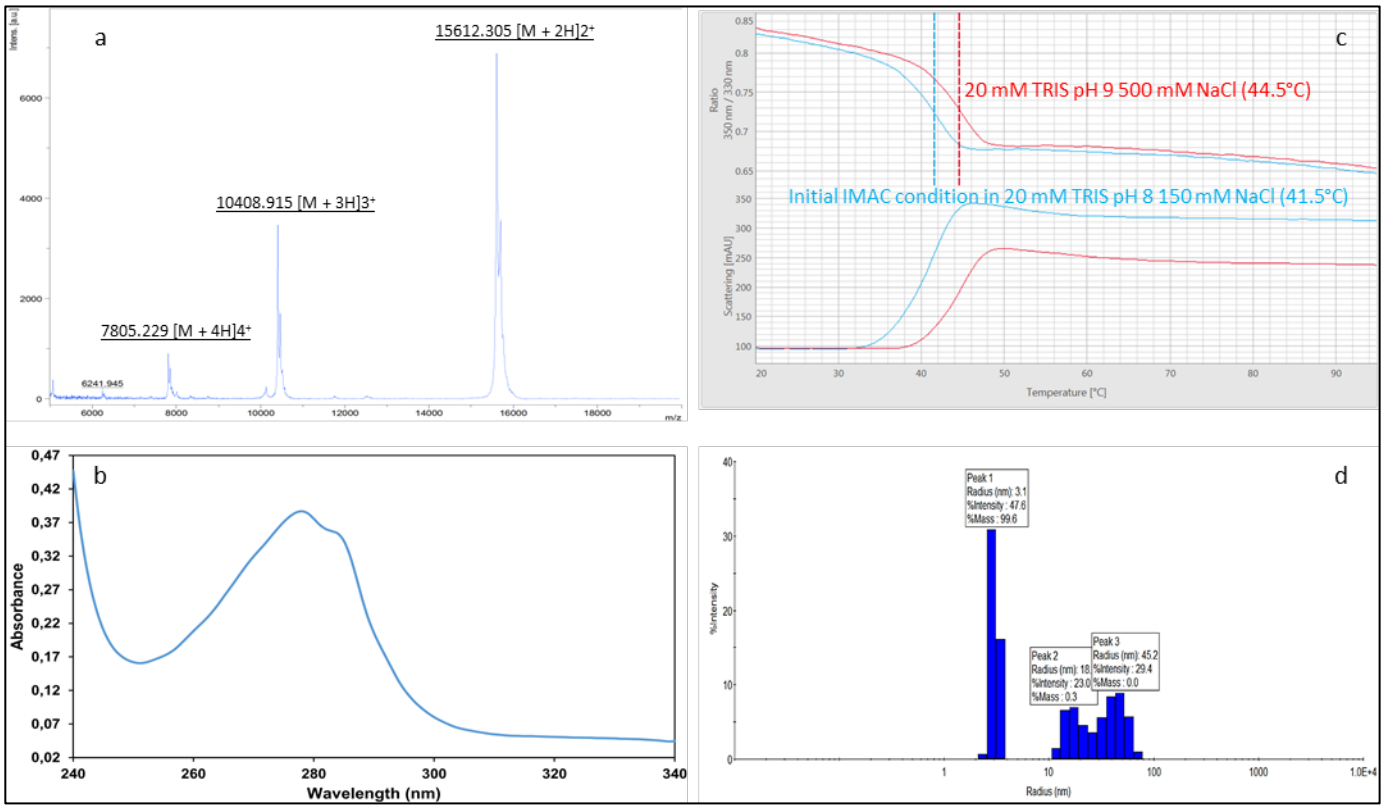
Supplementary Figure S1

SARS-CoV-2 SUD-M



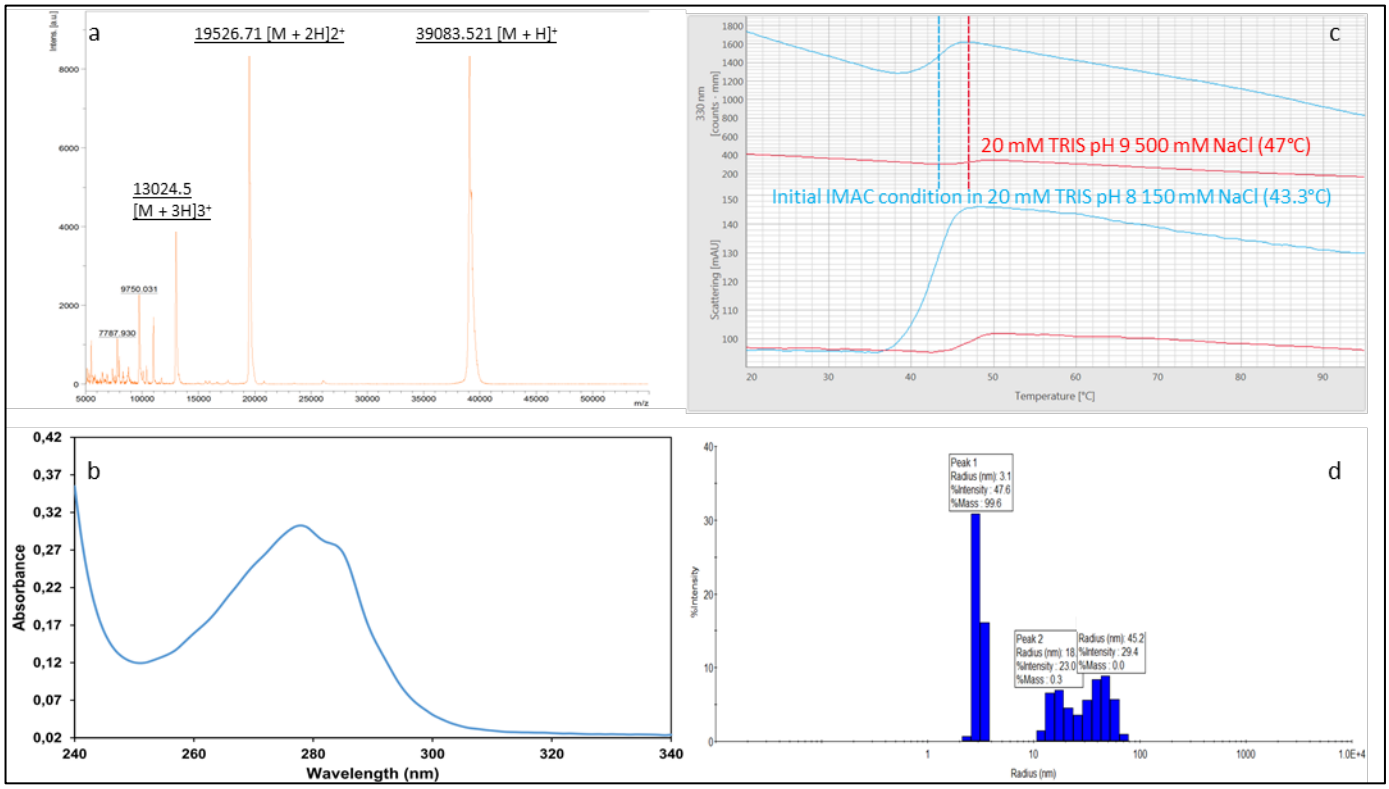
Supplementary Figure S2

SARS-CoV-2 SUD-NM

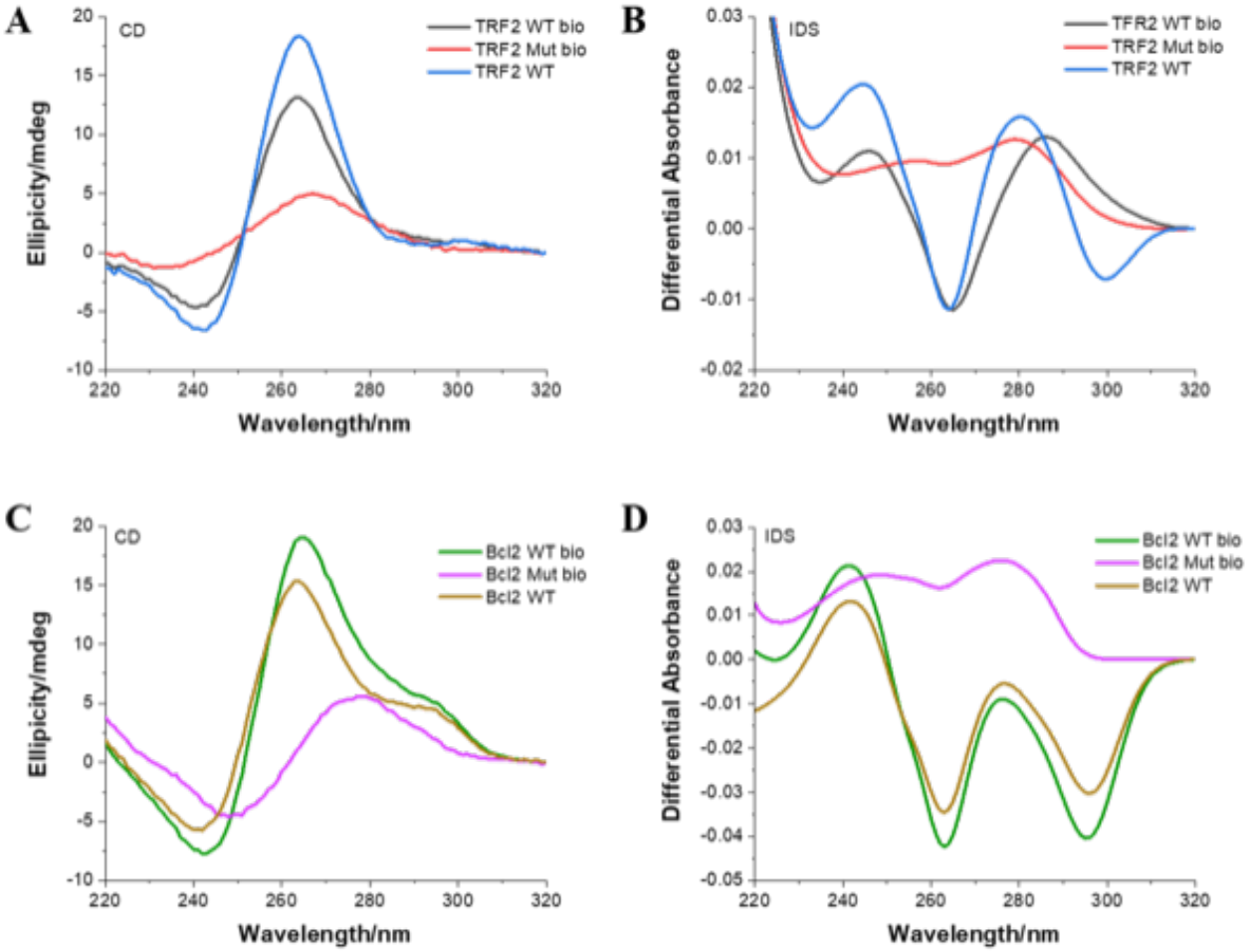


Supplementary Figure S3

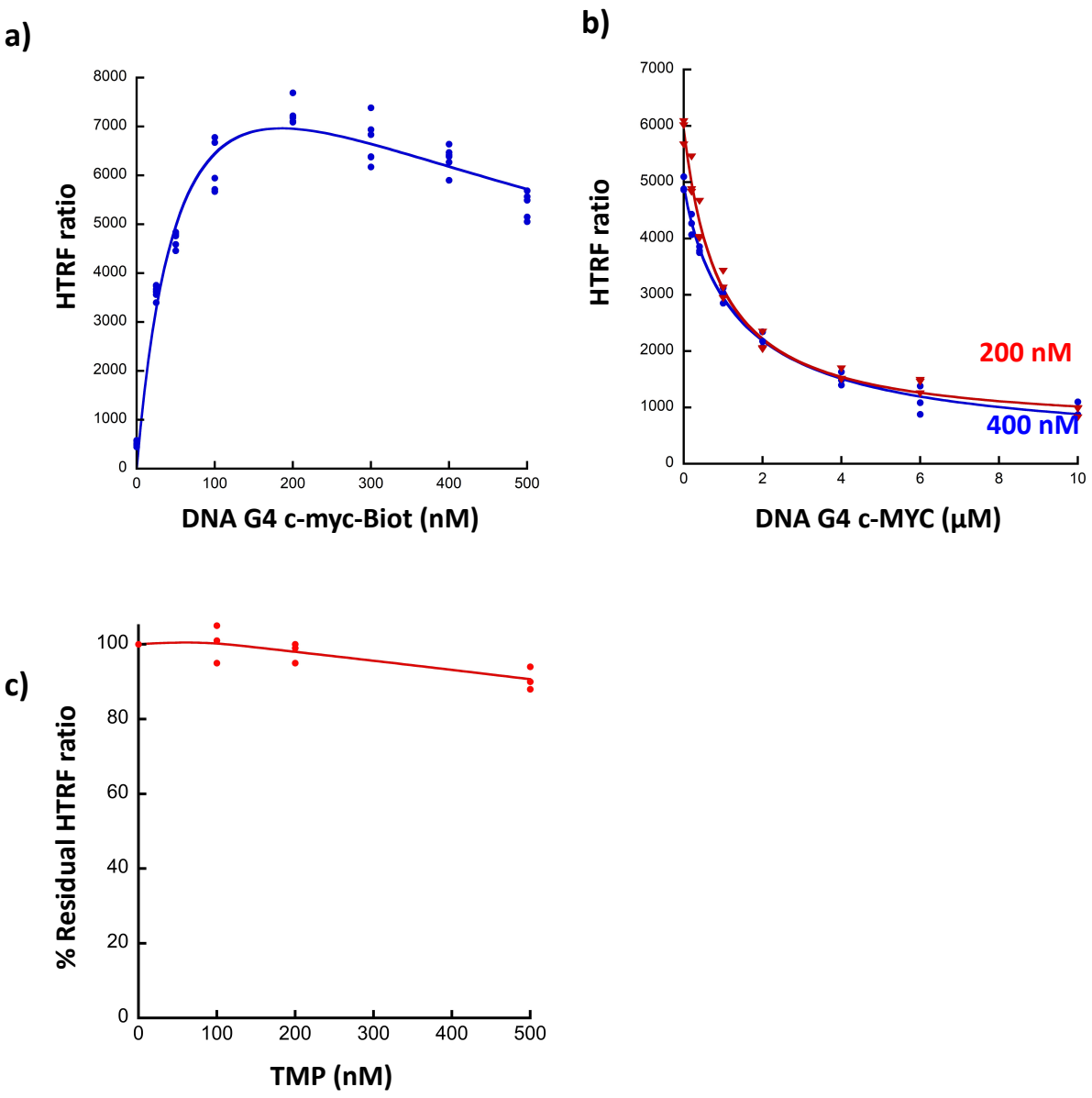
SARS-CoV-2 SUD-NMC



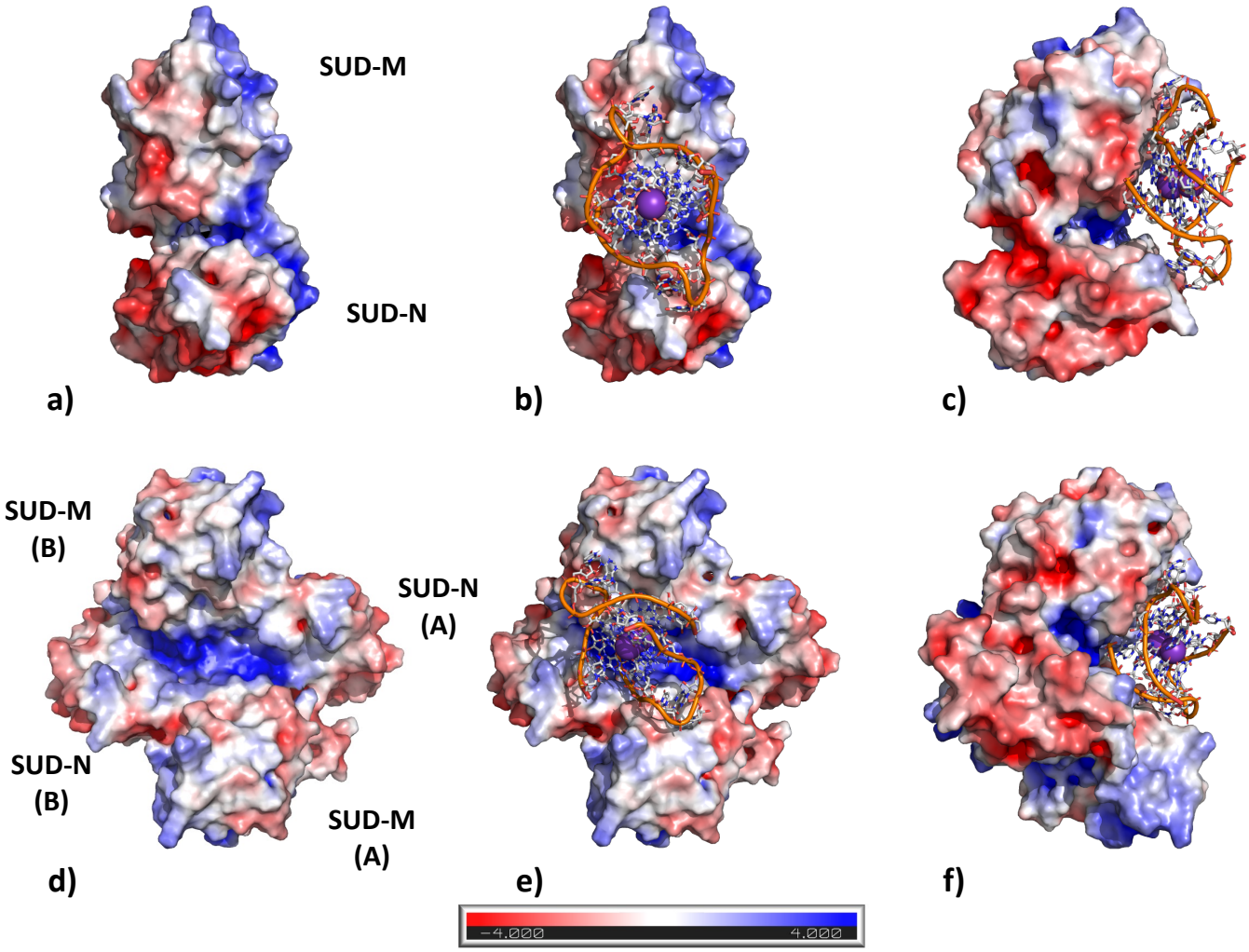
Supplementary Figure S4



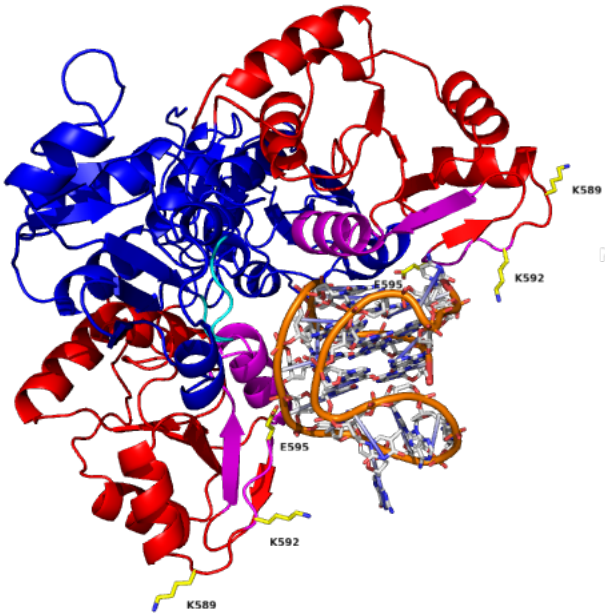
Supplementary Figure S5



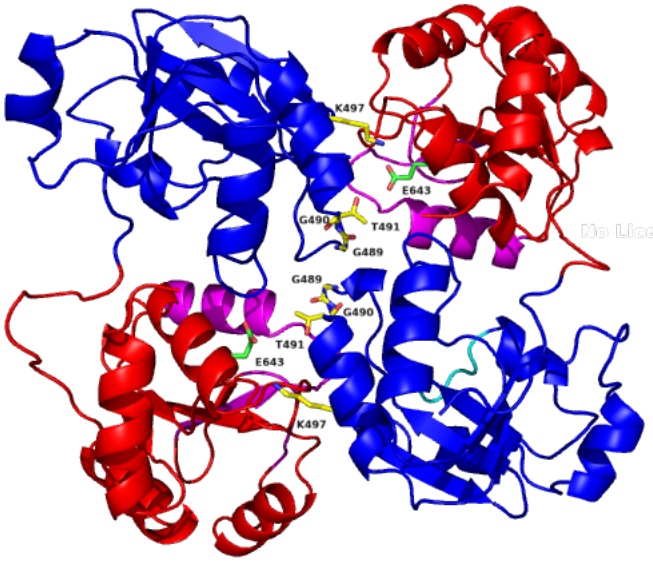
Supplementary Figure S6



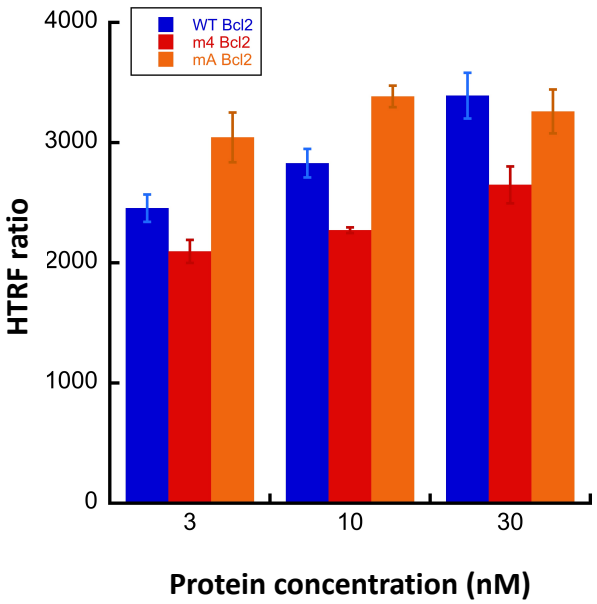
Supplementary Figure S7



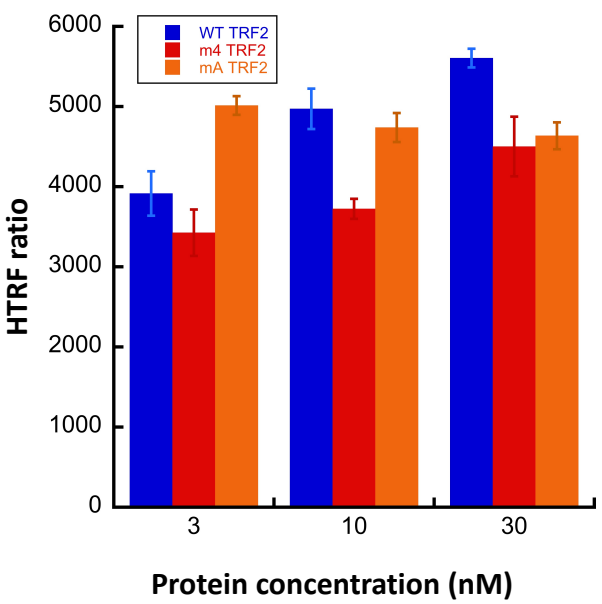
a



b

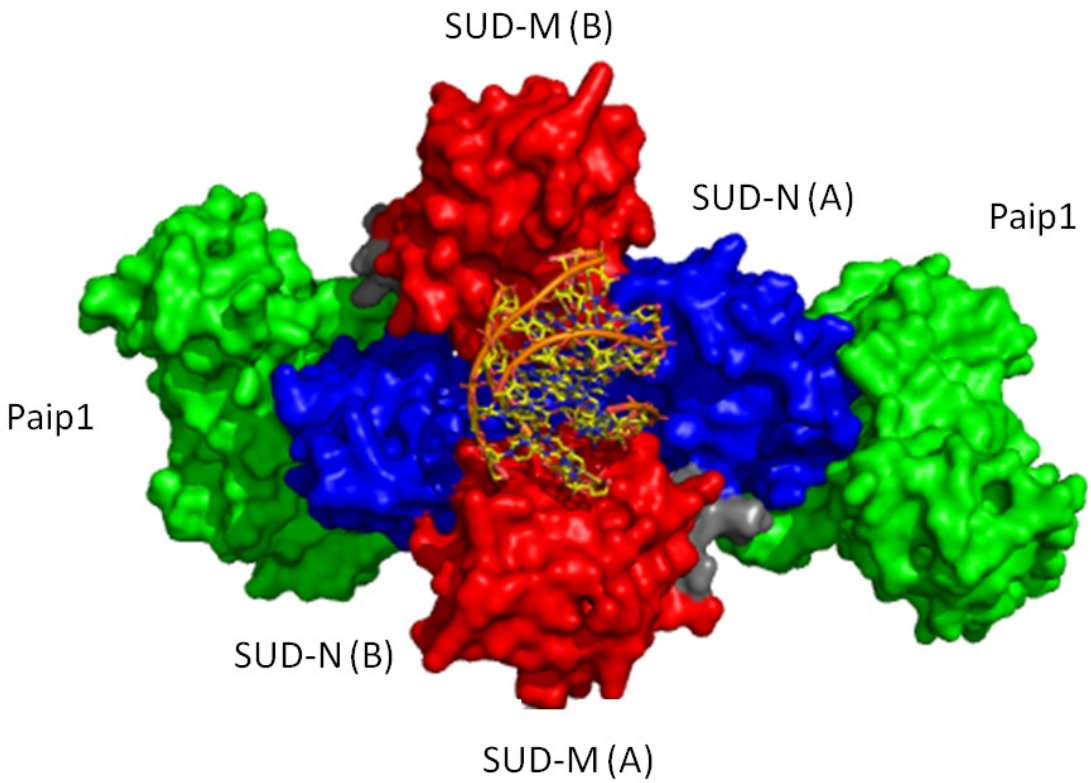


c



d

Supplementary Figure S8



Supplementary Figure S9