Electronic Supplementary Information

Microchip-based structure determination of low-molecular weight proteins using Cryo-Electron Microscopy

Michael A. Casasanta[a,b], G.M. Jonaid[a,c], Liam Kaylor[a,d], William Y. Luqiu[b,e], Maria J. Solares[a,d], Mariah L. Schroen[b], William J. Dearnaley[a,b], Jarad Wilson[f], Madeline J. Dukes[g], and Deborah F. Kelly[a,b*]

[a]Department of Biomedical Engineering, Pennsylvania State University, University Park, PA 16802, USA.

[b]Materials Research Institute, Pennsylvania State University, University Park, PA 16802, USA.

[c]Bioinformatics and Genomics Graduate Program, Huck Institutes of the Life Sciences, Pennsylvania State University, University Park, PA 16802, USA.

[d]Molecular, Cellular, and Integrative Biosciences Graduate Program, Huck Institutes of the Life Sciences, Pennsylvania State University, University Park, PA 16802, USA.

[e]Department of Electrical and Computer Engineering, Duke University, Durham, NC, 27708, USA.

[f]RayBiotech Life, Peachtree Corners, GA, 30092, USA.

[g]Applications Science, Protochips, Inc, Morrisville, NC, 27560, USA.

*Correspondence to: Debkelly@psu.edu

Supplementary Methods

N Protein validation. Purified N-protein (RayBiotech, 230-01104-100) was analyzed using denaturing (SDS-PAGE) or Native gels to confirm homogeneity. Gels were stained in SimplyBlue SafeStain overnight according and imaged using a BioRad ChemiDoc MP. The N protein (100 ng) was also analyzed via immunoblotting in parallel with SimplyBlue analysis to further confirm integrity of N-protein samples. Following electrophoresis, proteins were transferred to a PVDF membrane at 80 V in 1X NuPAGE transfer buffer for 60 minutes. The membranes were incubated in TBS-T supplemented with 3% BSA for 16 hours with gentle agitation. After blocking, membranes were incubated with antibodies raised against a 6x-His tag (GenScript, A00186) in TBS-T supplemented with 3% BSA at a dilution of 1:1000 at 25˚C for 60 minutes with gentle agitation. Following primary antibody incubation, membranes were washed with TBS-T with vigorous agitation. Membranes were incubated with α-mouse IgG-HRP conjugated secondary antibodies (1:10,000) in TBS-T for 60 minutes with gentle agitation. BioRad Clarity Max ECL blotting solution was used according to manufacturer's recommendations before visualization with a BioRad ChemiDoc MP.

RNA shift assays. 500 ng of His-tagged N-protein (0.92 mg/mL, RayBiotech, 230-01104-100) were incubated with 2.5 µg SARS-CoV-2 PCR+ (105 mg/mL; RayBiotech, CoV-PosPCR-S-100) human serum or SARS-CoV-2 PCR– human serum (103 mg/mL; RayBiotech, CoV-NegPCR-S-100) where indicated. Reactions were brought to a final volume of 20 µL with PBS. Reaction mixtures were then incubated at 25˚C or 37˚C for 60 minutes. After the indicated reaction time, sample loading buffer containing no SDS was added to each reaction mixture and samples were loaded into a NuPage Bis-Tris 4-12% polyacrylamide gel and allowed to migrate at 180 V for 60 minutes. After electrophoresis, proteins were transferred to a PVDF membrane using 1x NuPage transfer buffer at 80 V for 60 minutes. PVDF membranes were then incubated in TBS-T supplemented with 3% BSA at 4˚C with gentle agitation for 16 hours. Membranes were incubated with a primary antibody against 6x-His tag (GenScript, A00186) in TBS-T supplemented with 3% BSA at 25 ˚C for 60 minutes with gentle agitation. Membranes were washed with TBS-T and agitation then incubated with an α-mouse IgG HRP-

conjugated secondary antibody in TBS-T supplemented with 3% BSA for 60 minutes. Following secondary antibody incubation, membranes were washed with TBS-T for 10 minutes. Membranes were incubated with BioRad Clarity Max ECL substrate (1705062) according to manufacturer recommendations for visualization using a BioRad ChemiDoc MP.

Antibody binding assays. Aliquots (5 µL) of His-tagged N protein (0.1 mg/mL in 20 mM Tris (pH 7.5), 150 mM NaCl, 10 mM $MgCl_2$, 10 mM $CaCl_2$) were added to Ni-NTA-coated microchips (Protochips, Inc, EPT-52W) and incubated for 1 minute at room temperature. Control samples were incubated with Tris buffer solution that lacked the N protein. The excess solution was blotted away with Whatmann filter paper followed by the addition of either IgG+ (CoV-PosG-S-100) or IgG– (CoV-NegG-S-100) COVID-19 serum samples (0.3 mg/mL in 20 mM Tris (pH 7.5), 150 mM NaCl, 10 mM $MgCl_2$, 10 mM $CaCl_2$). Samples were incubated for 2 minutes at room temperature after which time the excess solution was blotted away with Whatmann filter paper. Chip contents were eluted with SDS-PAGE buffer solution (~10 µl per chip) and samples were assessed using 4-12% polyacrylamide gels and standard electrophoresis protocols. Gels were stained using SimplyBlue and visualized with a BioRad ChemiDoc MP.

Antibody detection kits. COVID-19 IgG rapid test kit cassettes (CG-CoV-IgG-RUO) were purchased from RayBiotech, Inc. along with IgG+ (CoV-PosG-S-100) and IgG– (CoV-NegG-S-100) serum samples. Aliquots of serum (25 µL) were mixed with the supplied sample buffer and applied to the sample area of the kit. Tests were read within 10 minutes of sample application. Three tests were run for each IgG+ or IgG– samples and no false positives were detected.

Supplementary Figures

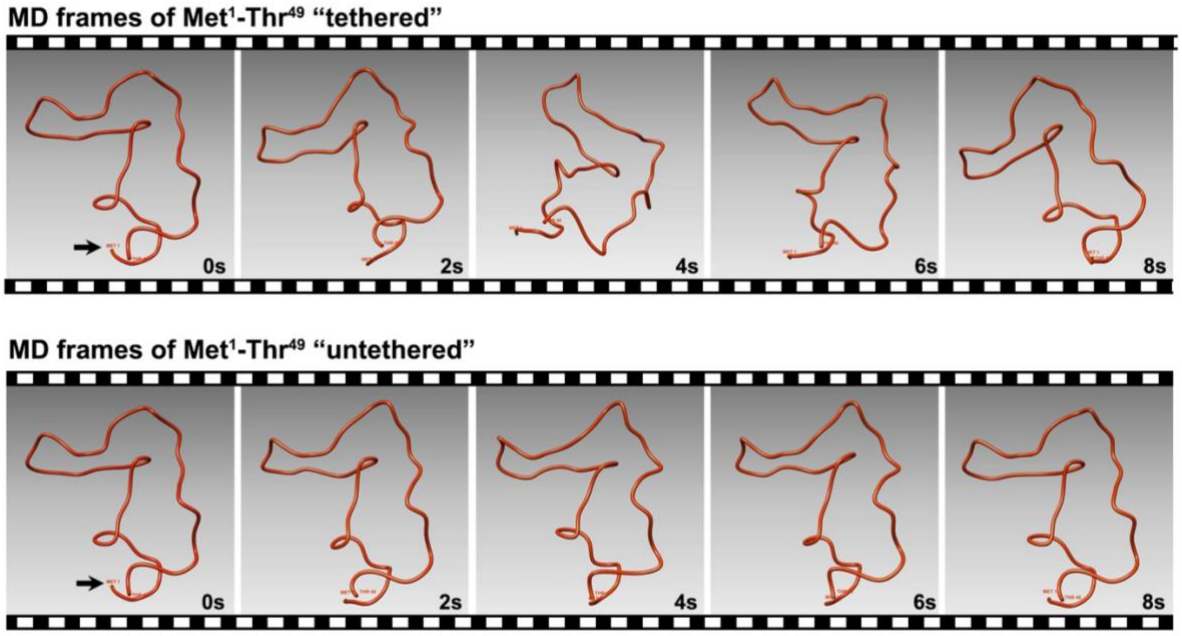

Figure S1. Molecular dynamics (MD) simulations show flexible loop domains in the N-terminal top hat region. The N-terminal residues of the N protein (Met^{1}-Thr^{49}) were examined using MD simulations integrated into the Chimera Software package[35]. Minimization parameters included 100 steepest descent steps with a step size of 0.02 Å, along with 10 conjugate gradient steps with a step size of 0.02 Å. Charges were assigned for minimization purposes using the AMBER force field (AMBER FF14SB). Simulations were performed on the initial structure for up to 300 frames of movie output. Met^{1} (black arrows) was either *(1)* held in place to represent the his-tagged "tethered" construct (top panel, **Movie S3**), or *(2)* allowed to move freely representing the "untethered" protein (bottom panel, **Movie S4**). Changes in the wire frame rendering of the protein segment show comparable differences in dynamic movements.

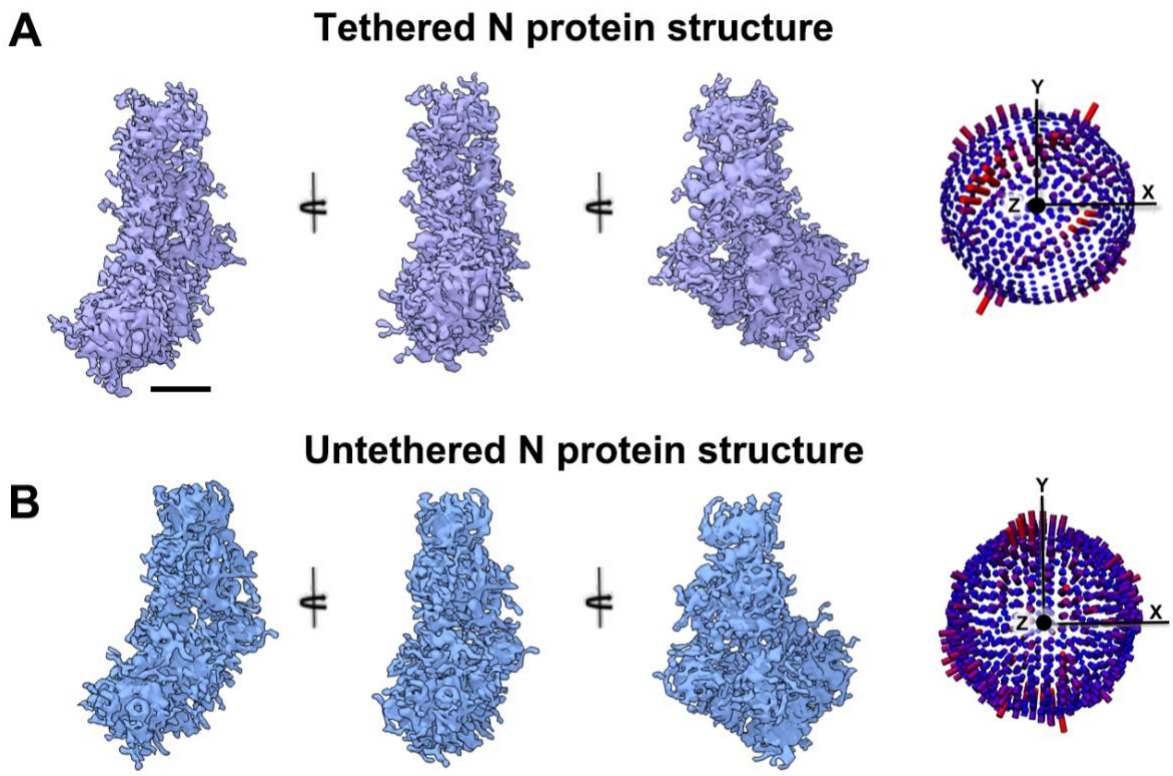

Figure S2. Comparison of microchip-tethered and untethered N protein structures. (A) EM structure of the his-tagged N protein sample that was tethered to the microchip substrate (purple) and resolved at 4.5-Å. **(B)** An untethered structure of the N protein (blue) interacted with microchip substrates through putative electrostatic charges (similar to glow-dicharged EM grids). The EM structures showed good agreement. Angular distribution plots of particle orientations (A, B) lacked major limitations and show a potentially greater number of particle views in the untethered structure. Scale bar is 10 Å.

Multi-sequence alignment for Pandemic- or CC-related coronavirus N proteins

Percent Identity to SARS-CoV-2 N-Protein						
SARS-CoV	Civet	MERS	OC43	HKU1	NL63	229E
89.74	89.50	48.59	35.9	35.66	27.86	25.22

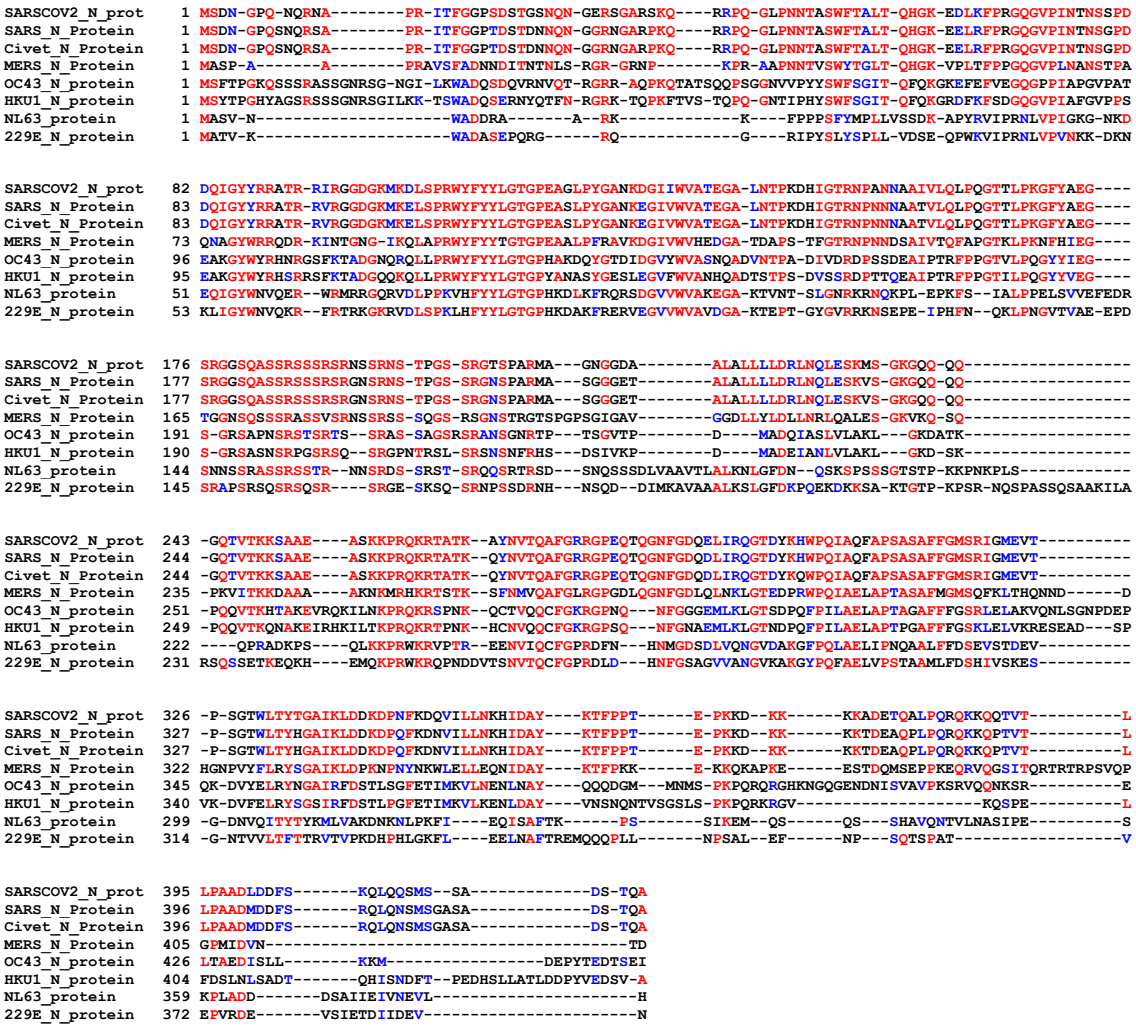

Figure S3. Multi-sequence alignment results for pandemic and CC coronavirus N proteins. Primary amino acid sequences for each N protein are listed in comparison to the SARS-CoV-2 pandemic strain. Sequences were obtained from ViPR and uploaded to the ClustalOmega online software package to generate percent identity matrices. Output shows identical amino acids in red and those with similar side chain properties are in blue.

Multi-sequence alignment for select CC-related coronavirus N proteins

Percent Identity to 229E N-Protein		
OC43	HKU1	NL63
28.21	29.8	45.38

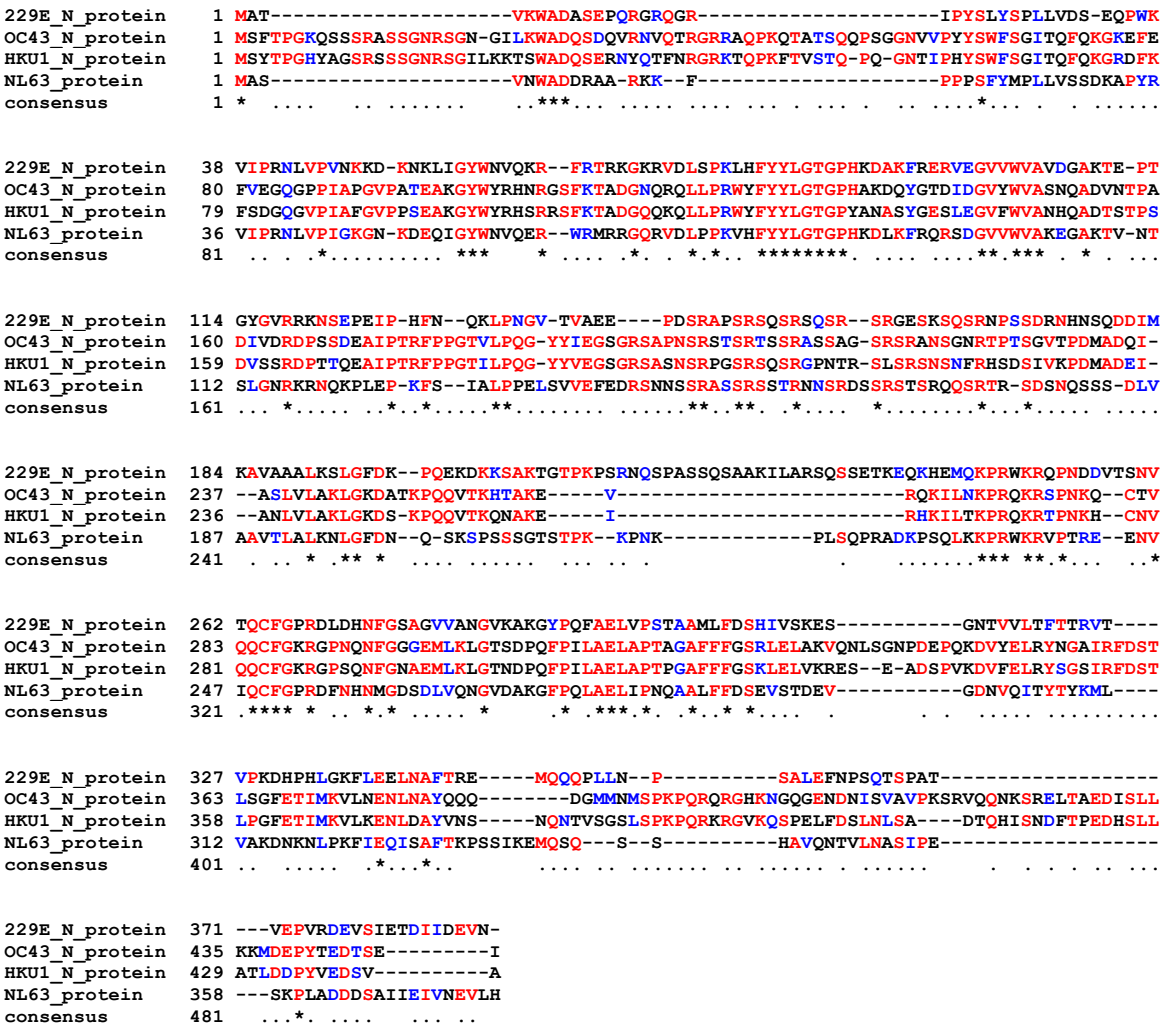

Figure S4. Multi-sequence alignment results for CC coronavirus N proteins. Primary amino acid sequences for each CC N protein were obtained from ViPR and are shown in comparison to 229E coronavirus strain. Sequences were uploaded to the ClustalOmega online software package to generate percent identity matrices.[48] Output is displayed with identical amino acids in red and those with similar side chain properties colored in blue.

Percent Identity to SARS-CoV-2 Predicted Epitope		
SARS-CoV	**Civet**	**MERS**
92.86	**92.86**	**33.33**

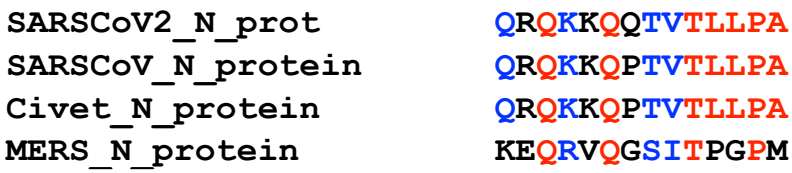

Figure S5. Multi-sequence alignment results for predicted epitope among pandemic strains. Primary amino acid sequences for the potential antibody binding site among pandemic strains are aligned for easy visual comparison. Percent identities were calculated using ClustalOmega and are recorded in the corresponding table.[48] Output is displayed with identical amino acids in red and those with similar side chain properties colored in blue.

Virus	GenBank ID
SARS-CoV-2	QJX60119.1
SARS-CoV	AAR86785.1
Civet-CoV	AAU04658.1
MERS	AKK52619.1
OC43	AXX83383.1
HKU1	ABG77571.1
NL63	ABI20791.1
229E	AAA45463.1

Table S1. Virus sequence information used in our structural analysis. A list of virus consensus sequences for pandemic and CC coronaviruses is given along with the GenBank ID number.

Supplementary Movies

Movie S1. Rotational views of the SARS-CoV-2 N protein map and model. The model for the SARS-CoV-2 structure (red) was placed into the EM map and shown in a variety of rotations. Areas of interest include the flexible N-terminal “top hat” motif that can adopt a variety of conformations and the more rigid C-terminal domain.

Movie S2. Sections through the SARS-CoV-2 N protein map and model. The N protein model (red) was placed into the EM map. The N- and C-terminal domains are labeled accordingly. Cross-sections through the structure are shown from two different views. The red model for the N protein disappears at the same rate as the surrounding density, which demonstrates a high-quality fit of the model within the map.

Movie S3. MD simulations show rotational freedom in the tethered N-terminal region. Dynamic simulations of a tethered N-terminal region (Met^{1}-Thr^{49}) demonstrate less internal movements than the untethered protein residues but long-range rigid-body rotational motion. Brownian motion was also observed in the protein domain, but in a more limited manner than the untethered protein segment (Movie S3). Simulations were performed on the initial structure for up to 300 frames of movie output.

Movie S4. MD simulations show flexible loops in an untethered N-terminal region. Dynamic motion in the flexible N-terminal residues (Met^{1}-Thr^{49}) demonstrate free, untethered movements. Long-range movements were limited although notable Brownian motion was observed. Changes in the wire frame of the protein fragment were examined and compared with simulations of a representative tethered domain (Movie S4). Simulations were performed on the initial structure for up to 300 frames of movie output.

Movie S5. Rotational views of the N protein bound to a Fab fragment. The N structure (red) and Fab fragment (cyan) were placed into the EM map and are shown in a variety of rotations. The epitope loop in the C-terminal domain of the N protein encompasses residues Q384 – A397. This loop is proximal to the Fab molecule and is seen protruding from the structure.

Movie S6. Sections through the N protein bound to a Fab fragment. Sections through the N protein (red) bound to a Fab domain (cyan) are shown from the side and top views. Models for the N-Fab complex disappear at the same rate as the surrounding density, indicating a high-quality fitting of the models within the EM density.

Movie S7. Comparison of N protein models from CC and pandemic-related coronavirus. NL43 (gray) and 229E (cyan) proteins belonging to the alpha-coronavirus family show similar overall features. HKU1 (blue) and OC43 (yellow) proteins that are beta-coronaviruses also show comparable architectures. These strains represent CC viruses. SARS (tan) and SARS-CoV-2 (red) models, while also classified as beta-coronaviruses, seem to adopt a variety of structural features seen in both alpha- and beta- strains. Structural features in the flexible N-terminal domain are unique to SARS-CoV-2 in comparison to the other models.

Movie S8. Comparison of N protein models from pandemic-related coronavirus. Models of the N proteins for SARS (tan) and MERS (pink) were aligned and have similar properties in the upper half of the protein with more variability in the lower half of the models. A comparison of N protein models for SARS-CoV-2 (red) and the Civet species (orange) show striking similarities in their flexible N-termini and more divergent features in the lower half of the protein, where more rigid features reside.