RESEARCH ARTICLE

Structural polymorphism of coiled-coils from the stalk domain of SARS-CoV-2 spike protein

Zala Živič[1] | Žiga Strmšek[2] | Marko Novinec[1] | Jurij Lah[1] | San Hadži[1,2]

[1]Faculty of Chemistry and Chemical Technology, University of Ljubljana, Ljubljana, Slovenia

[2]Department of Synthetic Biology and Immunology, National Institute of Chemistry, Ljubljana, Slovenia

Correspondence
Jurij Lah and San Hadži, Faculty of Chemistry and Chemical Technology, University of Ljubljana, Večna pot 113, Ljubljana, Slovenia.
Email: jurij.lah@fkkt.uni-lj.si and san.hadzi@fkkt.uni-lj.si

Abstract

Spike trimer plays a key role in SARS-CoV-2 infection and vaccine development. It consists of a globular head and a flexible stalk domain that anchors the protein into the viral membrane. While the head domain has been extensively studied, the properties of the adjoining stalk are poorly understood. Here, we characterize the coiled-coil formation and thermodynamic stability of the stalk domain and its segments. We find that the N-terminal segment of the stalk does not form coiled-coils and remains disordered in solution. The C-terminal stalk segment forms a trimeric coiled-coil in solution, which becomes significantly stabilized in the context of the full-length stalk. Its crystal structure reveals a novel antiparallel tetramer coiled-coil with an unusual combination of *a-d* and *e-a-d* hydrophobic core packing. Structural analysis shows that a subset of hydrophobic residues stabilizes different coiled-coil structures: trimer, tetramer, and heterohexamer, underscoring a highly polymorphic nature of the SARS-CoV-2 stalk sequence.

KEYWORDS
coiled-coil, SARS-CoV-2, spike, thermodynamics, X-ray
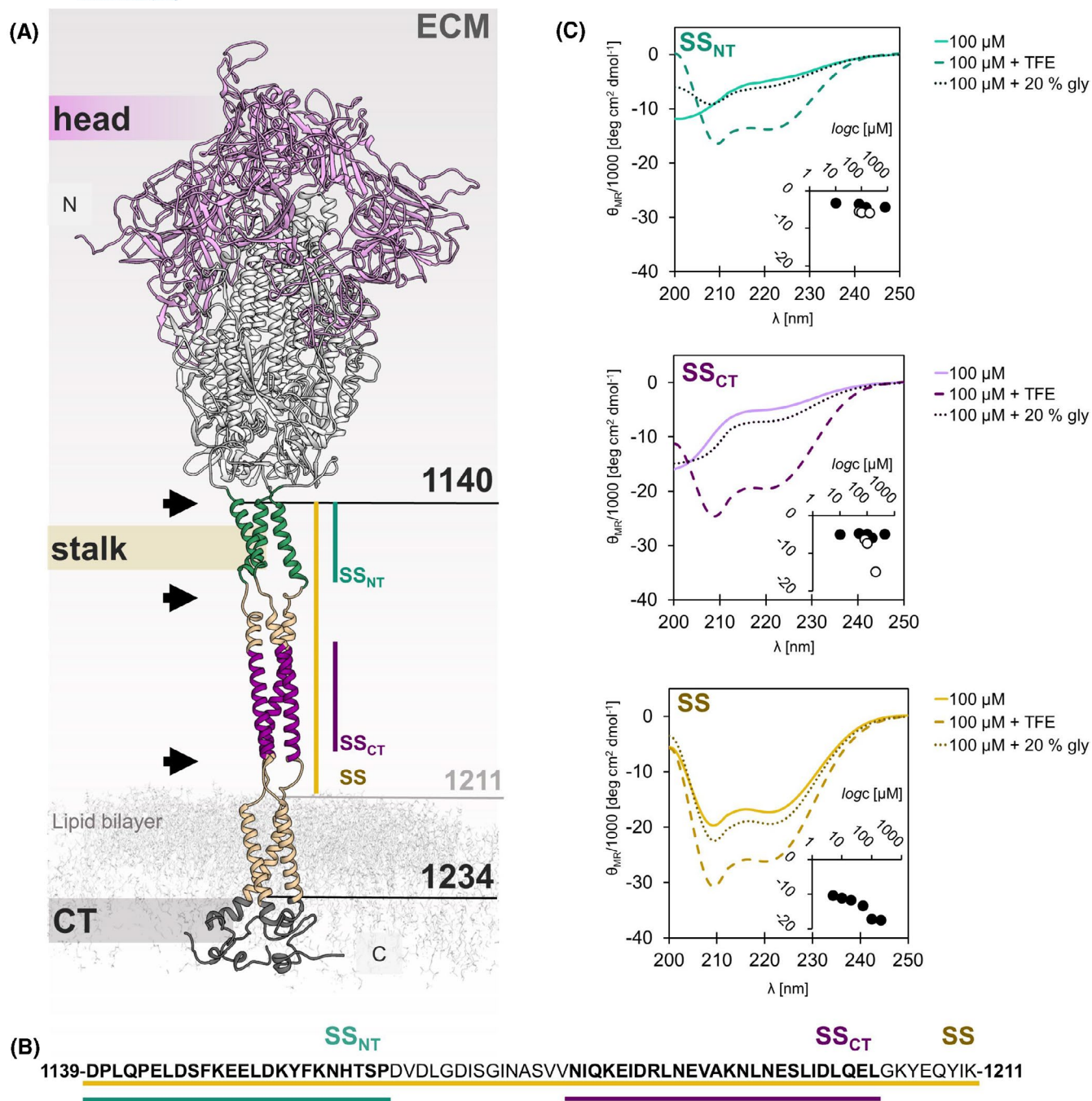
1 | INTRODUCTION

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a single-strand positive-sense RNA virus with four structural proteins (nucleocapsid, spike, membrane, and envelope protein) and 16 non-structural proteins.[1] Of these, the most extensively studied one is the spike protein due to its important role in the infection of host cells and its use in vaccines.[2–4] Spike protein is heavily glycosylated and forms trimers that are anchored to the viral membrane by a transmembrane helix.[5] A furin cleavage site in spike separates two functional subunits

S1 (residues 1-685) and S2 (686-1273). Structurally, spike can be divided into three distinct regions: head (1-1140), stalk (1141-1234), and the cytoplasmic tail (1235-1273) (Figure 1A).[6,7] Most of S1 and S2 are part of the head domain, whereas the stalk and the cytoplasmic tail represent the C-terminal part of S2. The binding of spike via the receptor-binding domain (330-530) to the host-cell receptor ACE2 is the initial step in the membrane fusion process that is followed by a dramatic conformational change of spike from its pre- to post-fusion conformation.[4] The pre-fusion conformation of spike is presumed to be a metastable structure with the head in the equilibrium

---

FIGURE 1 Formation of coiled-coils from the stalk domain of spike protein. (A) Different domains of spike protein: head (pink/white), stalk domain, and the cytoplasmic tail (CT, grey). S1 is colored pink. Different segments of the stalk domain are colored as tan (whole-length SS), green (N-terminal segment SS_{NT}), and magenta (C-terminal segment SS_{CT}). Black arrows indicate the location of three flexible hinges based on reference 10. (B) Stalk amino acid sequences used in this study are colored according to the scheme in the previous panel. (C) Circular dichroism spectra of SS_{NT}, SS_{CT}, and SS at 100 μM protein concentration in phosphate buffer or buffer supplemented with helix inducers (20% glycerol or 40% TFE). Insets show the value of MRE at 222 nm at different protein concentrations in buffer (black dots) or in buffer supplemented with 20% glycerol (white circles)

between "open" or "closed" conformation and a flexible stalk domain.[6] On the contrary, the post-fusion conformation takes on the form of a needle and is more stable and well-defined.[8]

While different pre-fusion conformations of the head have been extensively structurally characterized, the structural characteristics of the pre-fusion stalk domain remain poorly studied due to its high flexibility, rendering it invisible in the electron density maps obtained by cryo-EM. In one case, the N-terminal segment of the stalk adjacent to the head has been modeled as an unusual right-handed coiled-coil.[8] In the cryo-electron tomography

maps of intact viruses, the stalk domain is poorly resolved and a molecular model was built relying on the extensive simulation of molecular dynamics.[9,10] In this model, the stalk is highly helical, has two coiled-coil segments, and the flexibility is mediated by the three flexible hinges (Figure 1A).[7,10,11] Previous structural studies of isolated stalk domain and its segments from SARS-CoV, which has an identical amino acid sequence of stalk domain as SARS-CoV-2, have identified different coiled-coil assemblies, indicating a potential structural polymorphism.[12,13] Here, we use a combination of spectroscopy, small-angle X-ray scattering (SAXS), and protein crystallography to elucidate the stability and structural properties of the spike stalk domain.

2 | MATERIALS AND METHODS

2.1 | Computational sequence analysis

The sequence used to study spike stalk (SS) included residues 1139 to 1211 since residues 1139 and 1140 were observed to be a part of coiled-coil in one of the cryo-EM structures of the full-length spike.[8] Transmembrane region at the C-terminus was omitted. The end constructs SS (DPLQPELDSFKEELDKYFKNHTSPDVDLGDISGINA SVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIK) was analyzed with the online software Agadir using default settings to determine helix propensity and WAGGAWAGGA server that combines several coiled-coil prediction tools (Figures S1 and S2).[14,15] Two areas with significant helix propensity were selected as spike stalk N-terminal peptide (SS$_{NT}$, Figure S1/2, yellow) and spike stalk C-terminal peptide (SS$_{CT}$, Figure S1/2, green).

2.2 | Sample preparation

Peptides SS$_{NT}$ (DPLQPELDSFKEELDKYFKNHTSP) and SS$_{CT}$ (YNIQKEIDRLNEVAKNLNESLIDLQEL) were ordered from GenScript (Tokyo, Japan) and were >95% pure. Prior to use they were dissolved and dialyzed overnight against 20-mM phosphate buffer (10 mM NaH$_2$PO$_4$, 10-mM Na$_2$HPO$_4$, 150 mM NaCl, pH = 7). For in-house protein production, we constructed a modified spike stalk sequence containing a cleavage site for C-terminal Tobacco Etch Virus nuclear-inclusion-a endopeptidase (TEV protease) followed by His-tag (MDPLQPELDSFKEELDKYFKNHTSPDVDLGDISGINA SVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKE NLYFQGHHHHHH). The synthetic gene construct inserted into the pET-22b (+) expression vector was ordered from GenScript (Tokyo, Japan) and cloned into *Escherichia*

coli BL21[DE3] cells for protein expression. Cells were grown in LB media with 0.1 mg/ml ampicillin (Fischer Scientific, ≥98%, Thermo Fisher Scientific, MA, USA). When optical density at 600 nm reached 0.6 expression inductor isopropyl β-D-1-thiogalactopyranoside (IPTG, Merck, ≥99%, Darmstadt, Germany) at a final concentration of 0.5 mM was added. After overnight incubation at 20°C cells were then harvested for protein purification. Sonicated cells in phosphate buffer solution (1× PBS, 10 mM Na$_2$HPO$_4$, 1.8 mM KH$_2$PO$_4$, 137 mM NaCl, 2.7 mM KCl, pH = 7.4) were centrifuged for 30 min at 15 000 rcf. The filtered soluble phase was then used for further protein purification through nickel affinity chromatography with HisTrap high performance 1-mL column using fast protein liquid chromatography (FPLC) system Akta Pure 25 (GE Healthcare, IL, USA). His-tag removal was done with TEV protease cleavage. His-tagged TEV protease and His-Tag were removed with another nickel affinity chromatography using HisTrap high-performance 1-ml column, followed by a final purification step with gel chromatography using the Superdex 75 10/300 GL column (Cytiva, MA, USA) with Akta Pure 25 (GE Healthcare, IL, USA). Protein purity was assessed with SDS PAGE (Figure S3).

2.3 | Circular dichroism

All circular dichroism (CD) spectra were recorded with J-1500 Circular Dichroism Spectrophotometer (JASCO, MD, USA). Thermal denaturation spectra were obtained by measuring CD signals at 222 nm for temperatures from 5 to 80°C in intervals of 1°C at a speed of approximately 20°C/h. CD spectra at 25°C before and after denaturation were measured between 200 and 250 nm and were done in triplicates. Samples containing SS, SS$_{NT}$, or SS$_{CT}$ were diluted to different concentrations ranging from 5 to 1000 µM in phosphate buffer, in some cases supplemented with 40% TFE (TFE, Sigma–Aldrich, >99%, MO, USA) or 20% redistilled glycerol (Kemika, >99.5%, Zagreb, Croatia). Measurements were performed in 0.01, 0.05, 0.1, 0.2, and 0.5 cm quarts cuvettes. The obtained data (θ) was averaged and converted into mean residue ellipticity (MRE, θ$_{MR}$) given by Equation (1).

$$\theta_{MR} \left[\text{deg cm}^2 \text{dmol}^{-1} \right] = \frac{\theta * 0.1}{N_p * P_t * l} \tag{1}$$

N_p represents the number of amino acids in the sequence ($N_{SSNT} = 24$, $N_{SSCT} = 27$, $N_{SS} = 79$). P_t is the total monomer protein/peptide concentration and l is the optical pathway length. A two-state denaturation model was fitted to denaturation curves which is further described in the Supporting Information.

2.4 | Small-angle X-ray scattering

For small-angle X-ray scattering (SAXS) experiments, SS_{CT} and SS samples were prepared in phosphate buffer supplemented with 20% glycerol and PBS, respectively. They were measured on an Anton Paar SAXSpoint 5.0 equipped with Primux 100 micro Cu X-ray source, 2D 1M EIGER2 R series detector, heated/cooled sample cell holder, and ASX autosampler. Twenty microlitres of each sample at the concentration of 3.8 and 1.7 mg/ml for SS_{CT}, 3.9 mg/ml for SS, and their matching buffers were loaded into the 1 mm flow-through quartz capillary with the ASX autosampler. The temperature of the sample during measurements was kept at 10°C via a Peltier unit, while the waiting samples were kept at 4°C. 2D scattering patterns were collected as 10 frames of 30 min exposure per frame with a sample-to-detector distance of 900 mm. All measured scattering patterns were manually inspected for radiation damage. Appropriate 2D scattering patterns were analyzed and 1D raw SAXS curves were calculated with included software SAXS analysis (version 4.01). Further manipulation and analysis of SAXS curves were carried out using the ATSAS package.[16] BSA control, measured in the same set of experiments, was used to estimate the relative Mw along with other approaches as implemented in the ATSAS package (see Table S2).[16]

2.5 | SEC-R/LALS

Measurements were performed using OMNISEC RESOLVE (Malvern Panalytical, Malvern, United Kingdom) on a Superdex 200 Increase 10/300 GL column (Cytiva, MA, USA) in phosphate buffer at a flow rate of 0.5 ml/min at room temperature. 2 mg/ml BSA (Thermo Fisher Scientific, MA, USA) was injected three times at volumes of 100 μl as an internal standard before 200 μM SS was applied at 120 μl. The resulting data were analyzed using OMNISEC software version 11.20.

2.6 | Glutaraldehyde crosslinking

A fresh dilute of 2.3% glutaraldehyde (Merck, 50%, Darmstadt, Germany) was added to different concentrations of SS_{CT} in phosphate buffer with 20% glycerol at a ratio of 2.3% glutaraldehyde (V, [μl]): SS_{CT} (m, [μg]) = 1:10. SS_{CT} with added glutaraldehyde was incubated at 37°C for 2 to 5 min. Crosslinking was quenched by the addition of 1 μl of 1M Tris-HCl pH 8 followed by 10 min incubation at 37°C. SDS-PAGE of cross-linking products was performed using a Mini-PROTEAN Tris-Tricine gel (10%–20%, BIORAD, CA, USA) at a constant voltage of 100 V for 100 min. The maximum current was set at 65 mA. Before application to the gel, ~3 μg of protein with added 4× SDS loading buffer at a final volume of 12 μl was incubated at 70°C for 10 min. Images of destained gel were taken with ChemiDoc MP Imaging System (BIORAD, CA, USA).

2.7 | Crystallisation and structure determination

SS_{CT} was dissolved in Mili Q water at a concentration of 5 mg/ml and dispended on crystallization screens using Phenix robot (Art Robnins). Sitting drop crystallization plates were set by mixing 0.1 μl of protein with 0.1 μl of participant solution from different crystallization screens (PACT premier, Morpheus, JBScreen Wizard 3&4, SG1 Screen, The BCS Screen, JCSG Plus). After one week, crystals grew in 2-mM $ZnCl_2$, 0.1 M Tris, pH = 8, 20% (w/v) poly-ethylene-glycol (PEG) 6000. The crystals were flash-frozen in liquid nitrogen and cryo-protected using the crystallization solution with 30% glycerol.

X-ray data were measured at the SOLEIL synchrotron (Gif- sur-Yvette, France) at 100K on beamline PROXIMA1 using a PILATUS 6 M detector. Data were indexed, integrated, and scaled with XDS.[17] The phases were obtained using the direct method ARCHIMBALDO-LITE based on *ab initio* fragment search.[18] The structure was then manually rebuilt using Coot and refined using phenix. refine.[19,20] The final refinement cycles included TLS refinement (one TLS group per chain). Data collection and refinement statistics are given in Table S4.

3 | RESULTS

3.1 | Helix propensity of stalk domain and its fragments

We first investigated the coiled-coil and helix propensity of the stalk domain using bioinformatic tools. Helix prediction using AGADIR identifies two segments with a moderate helix propensity (Figure S1). These two helical segments coincide with the presumed N- and C-terminal coiled-coil segments of the stalk.[9,10] Different coiled-coil prediction algorithms, however, identify only the C-terminal segment as a potential coiled-coil, but not the N-terminal one (Figure S2). To elucidate the structural preferences of the stalk domain, we purified the stalk domain construct (SS, residues 1139-1211) and used synthetic peptides corresponding its N-and C-terminal segments (SS_{NT} and SS_{CT} covering residues 1139-1162 and 1178-1203, Figure 1B, Table S1). Circular dichroism (CD) spectra of these constructs show significant

differences in the helix propensity which increases from the almost unstructured SS_{NT} and SS_{CT} peptides (about 15% helicity) to the most helical SS (about 50% helicity, Figure 1C, Table S1). The addition of helix-stabilizing co-solutes glycerol or trifluoroethanol suggests that SS_{NT} has a lower helix propensity compared to the SS_{CT} peptide (Figure 1C, Table S1). To investigate whether these constructs form coiled-coils we recorded CD spectra at increasing protein concentrations since oligomerization is favored at higher protein concentrations and should be accompanied by an increase of CD signal due to the formation of coiled-coils. For the SS_{NT}, we did not observe any change in CD signal up to 800 μM concentration, neither in buffer nor in 20% glycerol, suggesting that the peptide does not oligomerize and form coiled-coils in line with the bioinformatic analysis (Figure 1C inset, Figures S1 and S2). For the SS_{CT}, we observed concentration-dependent changes in the CD signal only in presence of 20% glycerol, indicating that this segment has a moderate propensity for coiled-coil structure (Figure 1C inset, Figure S4).[21] The CD spectra of full-length stalk SS are strongly concentration-dependent with a transition midpoint around 100 μM protein concentration, indicating an even stronger propensity for coiled-coil formation (Figure 1C inset, Figure S4).

3.2 | Oligomerization state and thermodynamic stability

We next investigated the oligomerization state and thermodynamic stability of SS_{CT} and SS coiled-coils in solution. Initial SEC-R/LALS measurements of SS (Figure S5) revealed the presence of a minor and a major elution peak with the estimated molecular weight of 33 and 11 kDa, respectively. The molecular weight of the species in the major peak seems to be overestimated with respect to the theoretical weight of a monomer (9.1 kDa), however, we note that the ratio of molecular weight of two peaks suggests the presence of a trimer and a monomer. To clarify the oligomeric state, we performed SAXS measurements at concentrations where essentially all the samples should be oligomerized according to our CD data. The absolute scattering intensity with BSA as the reference indicates that the molecular weight of the SS is 27.2 ± 1 kDa, corresponding to the molecular weight of a trimer (Figure 2A, Table S2). Note that the molecular weight estimated from the absolute SAXS intensity is independent of molecular shape.

Due to the much smaller size of the SS_{CT} particles and buffer scattering the resulting SAXS measurements are noisier, nevertheless reliably indicating that the molecular weight is 9.5 ± 1 kDa also corresponding to a trimeric

state (Figure 2A, Table S2). We additionally verified the SS_{CT} oligomerization state using glutaraldehyde crosslinking. The separation of the crosslinking products on the SDS-PAGE shows a mixture of monomers and trimers as well as a small fraction of tetramers when using longer incubation times (Figure S6). For both SS and SS_{CT}, the thermal melting curves measured by CD show a cooperative concentration-dependent transition with a high degree of reversibility (Figure 2B, Figure S7). No such cooperative transitions can be observed for the SS_{NT} peptides, in line with other data showing that this peptide does not form coiled-coils (Figure S8). We analyzed the CD thermal melts using a two-state denaturation model assuming a transition from folded trimer to the unfolded monomer (Supporting Information Equations S1–S4). Global fitting of the model function to the temperature and concentration-dependent CD data series simultaneously yielded a set of thermodynamic parameters describing the coiled-coil stability, which are reported at the reference temperature $T_0 = 25°C$ in phosphate buffer for SS and in 20% glycerol buffer for SS_{CT} (Table S3). The denaturation free energies $\Delta G_{D,T0}$ are 11.6 and 7.5 kcal mol^{-1} for SS and SS_{CT}, respectively. The higher stability of the SS compared to SS_{CT} may indicate that the coiled-coil segment is longer in the context of SS; however, additional stabilization comes also from the larger size of the construct, since helix folding is more favorable for longer peptide chains.[22] While SS is more stable compared to SS_{CT}, the unfolding of both proteins is accompanied by similar values of enthalpy and heat capacity change ($\Delta H_{D,T0} = 23$ and 19 kcal mol^{-1} and $\Delta c_{p,D} = 1$ and 1.3 kcal mol^{-1} K^{-1}, Table S3). Given that the magnitude of heat capacity change correlates with the protein surface area exposed upon denaturation,[23] the similar values of $\Delta c_{p,D}$ suggest that the packing of SS_{CT} coiled-coil core likely presents the majority of surface-area change in the SS domain. The somewhat higher $\Delta c_{p,D}$ for SS could be due to the elongation of the SS_{CT} coiled-coil in the context of SS, resulting in a larger surface surface-area exposure upon denaturation.

3.3 | Structure of SS_{CT} coiled-coil tetramer reveals unusual core packing

To delineate the coiled-coil structure of SS_{CT} at the atomic resolution we used X-ray crystallography. The crystals of SS_{CT} peptide diffracted to 1.9 Å resolution and belong to space group $P2_12_12_1$ (see Table S4 for data collection and refinement statistics). Given the natural trimeric state of the spike and the properties of SS_{CT} in solution, we expected the peptide to crystalize in a trimeric state. However, the structure reveals a single tetramer of antiparallel coiled-coils in an asymmetric unit (Figure 3A).
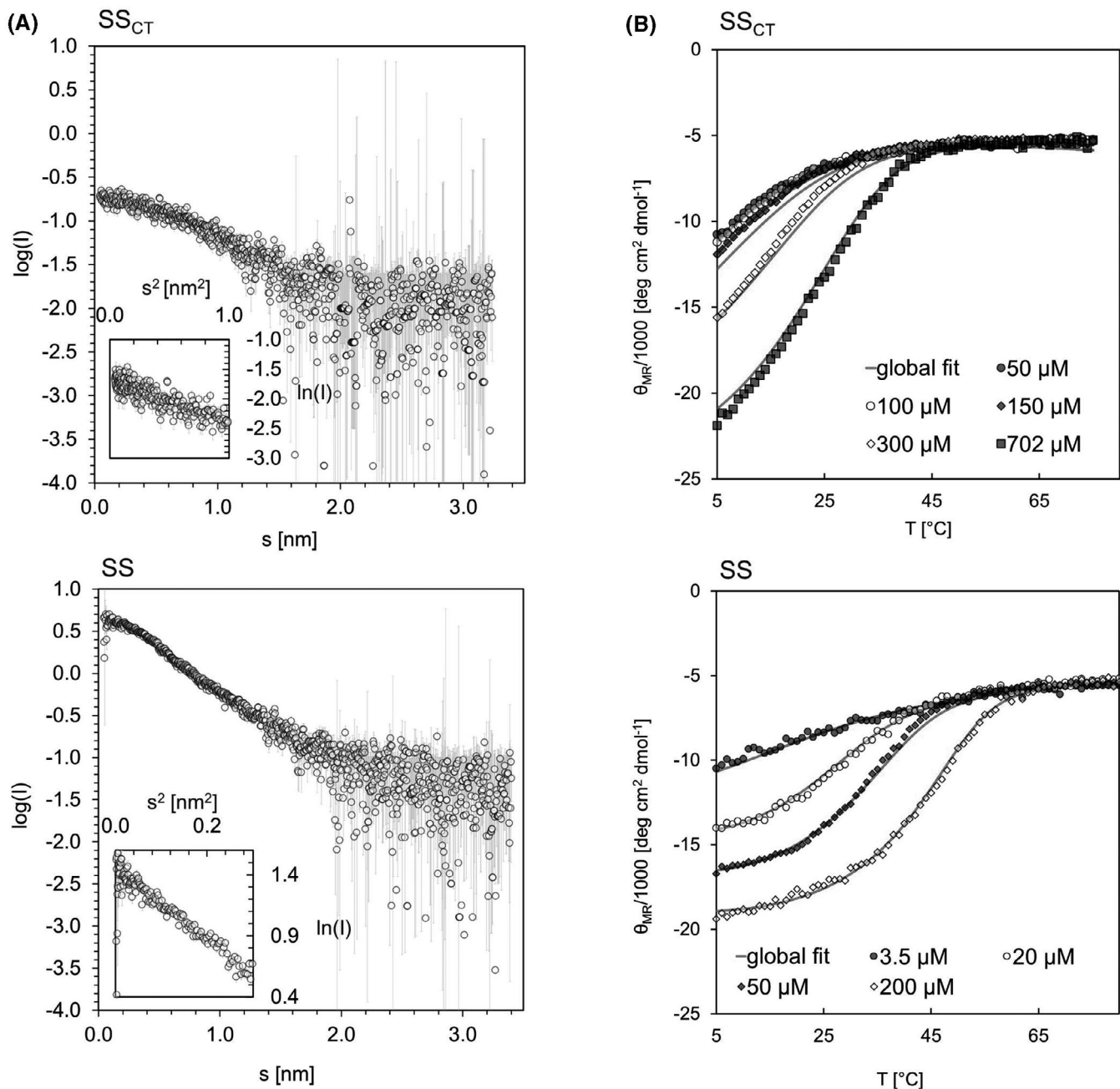
FIGURE 2 Solution properties of the SS and SS_{CT} coiled-coils. (A) SAXS profiles for SS_{CT} (top) and SS (bottom) were measured in the phosphate buffer with 20% glycerol or buffer only, respectively. Inset shows the Guinier region. The molecular weight estimated using different approaches is listed in Table S2. (B) Thermal denaturation of SS_{CT} (top) and SS (bottom). Mean-residue ellipticity at 222 nm is shown as symbols for different protein concentrations. A global fit of the two-state model (Equation S1) is shown as a gray solid line with the parameters reported in Table S3

Overall, the complex has a cylindrical shape, is about 38 Å long, and buries 5000 Å^2 of solvent-accessible area (Figure 3A).

Adjacent helices run antiparallel to one another and stand at an angle of ~20°, while the diagonally related helices have the same relative orientation. The tetramer hydrophobic interface is stabilized by a non-classical pattern of core packing interactions. Two helix cross-sectional *a-d* layers in the middle of the tetramer form a canonical *a-d*

core (layers 1 Figure 3B). In these layers, the hydrophobic residues Val1189 at position *d* and Leu1193 at position *a* interact via knob-into-holes packing and form a 4-knob cycle (Figure 3B). In contrast, the cross-sectional *a-d* layers toward helix termini (layers 2 and 3 in Figure 3A) are arranged via non-canonical *e-a-d* core packing. In layer 2, two Leu1186 at position *a* from the diagonally related helices face each other (knobs-to-knobs packing) and point to the center of the tetramer core. On both sides they are
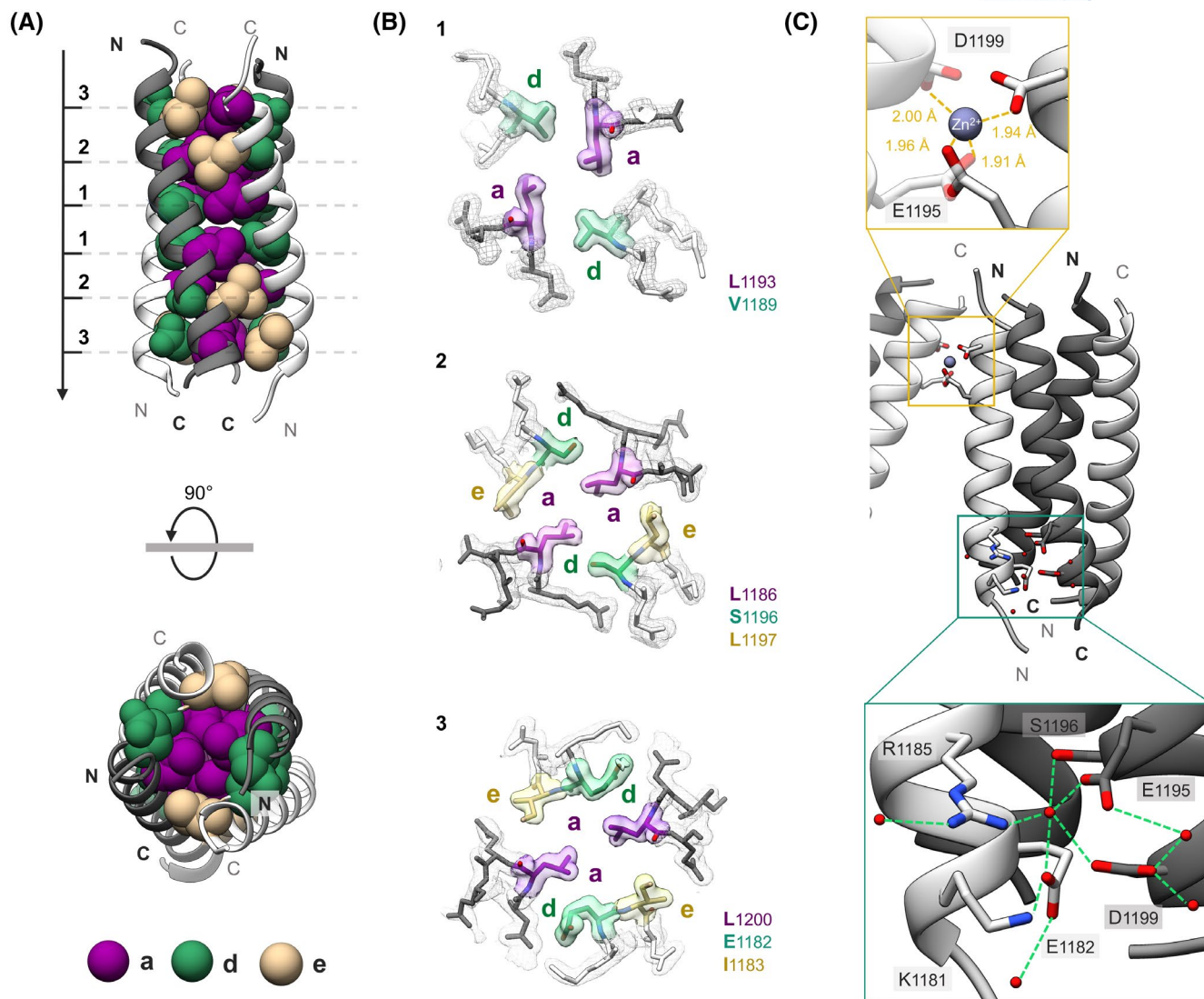
FIGURE 3 Crystal structure of the antiparallel coiled-coil tetramer of the stalk segment SS_{CT}. (A) The overall structure of tetramer. A pair of helices with the same orientation is depicted as white or grey. Amino acids at positions *a*, *d* and *e* are shown as spheres (dark magenta, sea green). Helix cross-sectional *a-d* layers are labeled 1, 2, and 3. (B) The 2Fo-Fc electron density map (mesh, contoured at 1.0 σ) is shown with the refined molecular model of the three *a-d* layers (sticks). (C) A network of electrostatic interactions stabilizes pairs of antiparallel chains (bottom panel). One of these networks is disrupted by the Zn^{2+} ion that is chelated by a pair of negatively charged sidechains from a symmetry-related tetramer (upper panel)

shielded by the Ser1196 and Leu1197 at positions *d* and *e*, forming an extended hydrophobic core (Figure 3B). Similarly, layer 3 Leu1200 packs knobs-to-knobs and is shielded by Glu1182 and Ile1183 at positions *d* and *e*. This unusual combination of both *a-d* and *e-a-d* core packing is reflected in the progressive changes of the relative surface burial along the sequence (Table S5). The residues at position *e* change from being exposed in the middle layer 1 with the *a-d* core and become more buried in the terminal layers with the *e-a-d* core packing. Formation of this type of non-canonical packing requires axial rotation of the individual helices relative to the tetramer axis, which shifts *a* or *d* position toward the center of the tetramer.[24] Indeed, we observe that the helices A and D are rotated by around

−22°, while the axial rotation of helices B and C changes along the sequence and increases from −22° to −14° in the layers with the canonical *a-d* packing (Figure S9). Apart from the hydrophobic core packing, two pairs of chains running antiparallel to each other (A–B and C–D) are additionally stabilized by a network of electrostatic interactions between Lys1181 and Arg1185 at positions *c* and *g* and the oppositely charged Glu and Asp on the neighboring chain (Figure 3C). The Arg1185 sidechain packs into holes formed by these charged residues and in some arrangements form short-distance (<3 Å) salt bridges. Interestingly, one of these networks is disrupted by a Zn^{2+} ion, which is coordinated by the pair of Glu and Asp sidechains from the symmetry-related tetramer, essentially

forming a Zn^{2+}-mediated crystal contact (Figure 3C). Collectively, the structure of SS_{CT} reveals a novel combination of *a-d* and *e-a-d* core packing stabilizing an antiparallel coiled-coil tetramer.

4 | DISCUSSION

Here, we investigated the coiled-coil formation of peptides derived from the stalk domain of the SARS-CoV-2 spike protein. The isolated N-terminal SS_{NT} stalk segment showed no tendency to form coiled-coil in solution. In one of the structures of the spike protein reconstructed from the single-particle cryo-EM this segment has been modeled as a right-handed parallel coiled-coil trimer with a loosely packed core formed by phenylalanine sidechains (Figure S10).[8] In other spike models the SS_{NT} segment however appears as the left-handed coiled-coil trimer.[7] While it is possible that such structure exists in the context of the whole spike protein where it is additionally stabilized by the trimerization of the head domain, our data indicate that the SS_{NT} coiled-coil is not stable in isolation. In contrast, the whole length stalk SS construct as well as its C-terminal segment SS_{CT} do form coiled-coil trimers in solution. The NMR model of a shorter SARS-CoV stalk segment as used in our study (residues 1159-1211) but with the identical amino acid sequence also shows a parallel coiled-coil trimer (Figure 4A). This trimer is stabilized by *a-d* packing of leucine, isoleucine, and alanine sidechains, but lacks favorable electrostatic interactions between the *e* and *g* positions (Figure 4A). Based on the sequence similarity and our results, we assume that the SS and SS_{CT} constructs form such parallel trimers in the solution. The calculated helicity from the CD spectra of SS indicate that around 36 residues are helical, which corresponds to the helicity observed in the mentioned NMR model. This then indicates that the SS_{NT} segment does not become additionally structured in the context of SS construct, but the gain in helicity is likely due to elongation of the SS_{CT} coiled-coil by one heptad repeat, which would result in the structure as observed by NMR. This hypothesis is in accordance with the observed slightly larger heat capacity change of SS compared to SS_{CT} (Table S3) since the formation of additional SS_{NT} coiled-coil in the SS context should give a significantly larger difference in SS and SS_{CT} heat capacities.

Additionally, it is interesting to note the discrepancy of some spike stalk properties and their models. In two previously mentioned models of the whole spike, the stalk region has been modeled either entirely helical and then subjected to molecular dynamics studies[10] or is considered to be about 77% helical[7] which is only roughly achieved (72%) by the addition of TFE to high concentrations (Table S1). In buffer solution, the observed helicity is about ~50%, which likely corresponds to the helicity of SS_{CT} elongated by one heptad. Clearly, the full-length protein might contribute to the stabilization of additional helical or coiled-coil structures (e.g., in the N-terminal part); however, it is important to consider the possibility of a more disordered stalk region when modeling the protein.

Whereas in solution SS_{CT} forms trimers, it adopts an antiparallel tetramer conformation in the crystal form. This has also been observed for a similar, 17 amino acids longer SARS-CoV peptide (termed C44) with the otherwise identical sequence. The structures of these two tetramers, however, importantly differ in several aspects. Foremost in the SS_{CT} tetramer all helices are in register, while in the C44 tetramer a pair of chains with the same orientation in are out of register by about one helix turn relative to the second pair (compare Figure 4B,C). Furthermore, the C44 tetramer is stabilized by the non-canonical *a-d-g* core packing where Leu and Ile sidechains at *d* positions pack knob-to-knobs and are shielded by *a* and *g* residues from the neighboring helix (Figure 4E). On the contrary, we observe a combination of *a-d* and *e-a-d* core packing in the SS_{CT} tetramer.

Finally, the SS_{CT} stalk segment also importantly contributes to the stabilization of heterohexamer in the postfusion spike conformation (Figure 4D). Here, the inner trimeric core is formed by the heptad repeat 1 (HR1), while the stalk domain packs between three chains in the antiparallel orientation. The SS_{NT} segment adopts an extended conformation stabilized by the Asn/Gln zippers,[25] while the SS_{CT} segment remains helical and contributes hydrophobic sidechains to stabilize the central hydrophobic HR1 cluster. Again, we observe that similar residues as those in trimer and tetramer conformations are involved in the stabilization of this structure. To understand the energetic basis of individual residues in the stabilization of different stalk structures (trimer, two tetramers, and heterohexamer) we performed a computational alanine scanning (Figure S11). In accordance with our structural observation, we find that the majority of residues mediating energetically most important interactions are shared between all four conformations of the stalk. These hotspot residues are mainly hydrophobic, only in two structures charged residues make important energetic contributions (Figure S11). Lack of more specific electrostatic interactions between the *e* and *g* positions and predominance of hydrophobic ones is the likely reason for the high structural polymorphism of this protein sequence. Structural polymorphism, where a given sequence does not uniquely define the structure, is often encountered in the field of coiled-coil design, where the precise control of orientation or oligomeric state can only be achieved using negative design.[26,27]
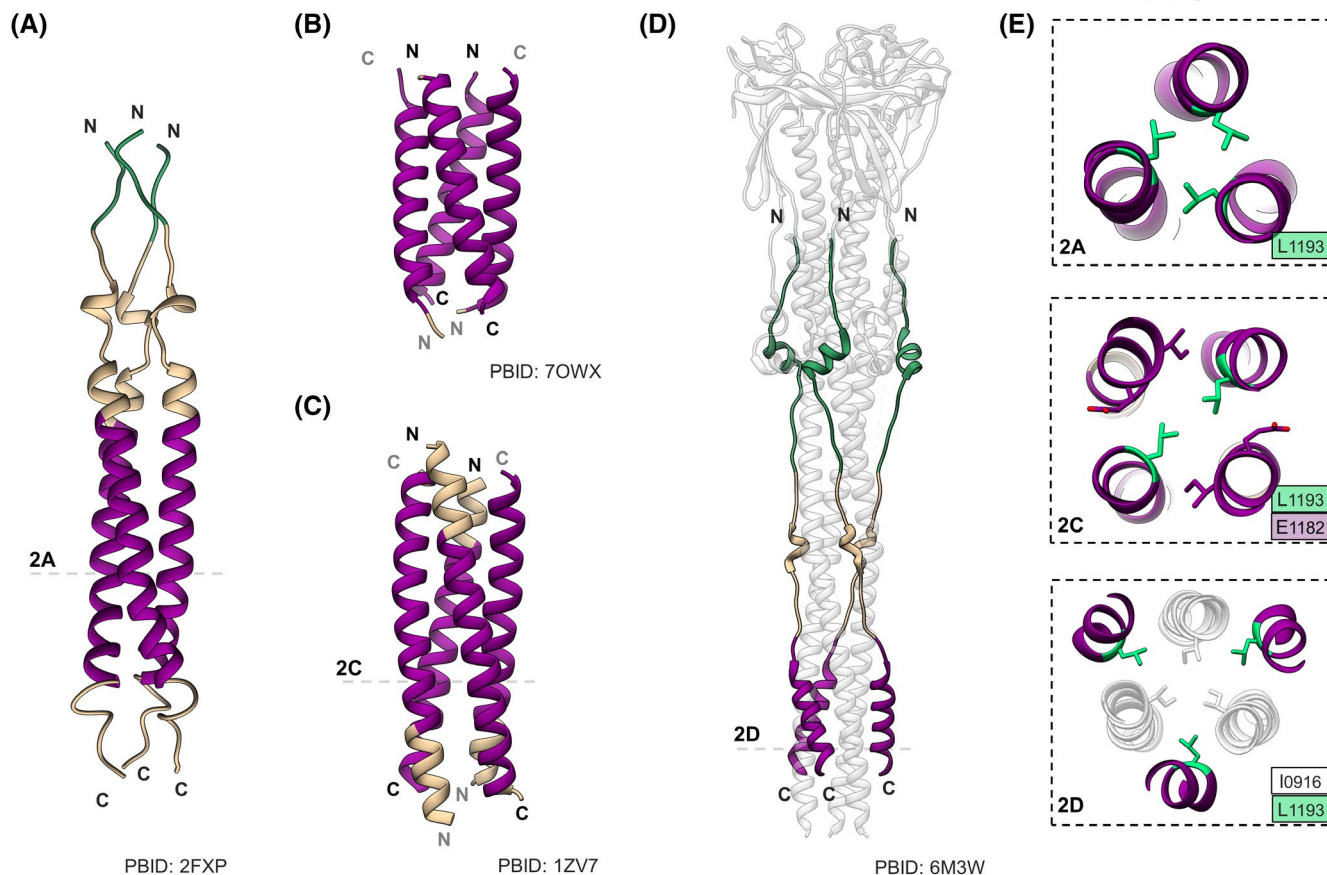
FIGURE 4 Structural polymorphism of the stalk domain. Residues from the SS_{CT} sequence are colored in magenta and adopt various conformations depending on the structural context. Residues belonging to SS_{NT} are shown in green. (A) The NMR model of the pre-fusion parallel coiled-coil trimer (residues 1159-1211). (B) Crystal structure of the antiparallel coiled-coil tetramer SS_{CT} (also shown in Figure 3). (C) Crystal structure of the antiparallel coiled-coil tetramer C44 (residues 1168-1211). (D) Structure of the post-fusion heterohexamer (residues 706-1196) is shown as a transparent model, stalk region is shown in colors. (E) Cross-sectional layers centered around residue L1193 show the differences in the hydrophobic packing between these structures. The corresponding layer for the SS_{CT} tetramer is shown in Figure 3B (layer 1)

Collectively, the new coiled-coil structure of the SS_{CT} tetramer underscores the highly polymorphic nature of the stalk segment, meaning that a single sequence can encode different structures. The model of membrane fusion of class I fusion proteins postulates that the metastable pre-fusion conformation transitions into a thermodynamically much more stable post-fusion conformation, which provides a reservoir of free energy required for the refolding of the spike.[28] It has been suggested previously, that different structures of stalk segments including C44 tetramer from the SARS-CoV peptides might represent intermediate states of S2 during the fusion process.[12] While the evidence for this hypothesis is still lacking, it is clear that the stalk sequence is highly polymorphic since a small subset of hydrophobic residues can stabilize different final structures (Figures 4 and S11). In other words, the observed polymorphism of the stalk domain seems to reflect a rather flat and monotonous free energy landscape for this protein sequence, where several conformations are separated by a small free energy difference, keeping the chain in the meta-stable conformation and priming it for refolding into the energetically more favorable post-fusion state.

DISCLOSURES

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Zala Živič performed experiments and wrote the manuscript. Žiga Strmšek Measured and analyzed SAXS data. Jurij Lah and Marko Novinec discussed the results and helped with data analysis. San Hadži conceived the study, coordinated the experiments, and wrote the manuscript.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available in the methods and Supporting Information of this article. Coordinates and structure factors have been deposited in the Protein Data Bank under the ID code 7OWX for SS_{CT}.

ORCID

Zala Živič https://orcid.org/0000-0001-9044-3064
San Hadži https://orcid.org/0000-0002-5615-7918

REFERENCES

1. Wang M, Zhao R, Gao L, Gao X, Wang D, Gallagher T. SARS-CoV-2: structure, biology, and structure-based therapeutics development. *Front Cell Infect Microbiol.* 2020;10(November):1-17. doi:10.3389/fcimb.2020.587269

2. Huang Y, Yang C, Xu X, Xu W, Liu S. Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19. *Acta Pharmacol Sin.* 2020;41(9):1141-1149. doi:10.1038/s41401-020-0485-4

3. Krammer F. SARS-CoV-2 vaccines in development. *Nature.* 2020;586(7830):516-527. doi:10.1038/s41586-020-2798-3

4. Hoffmann M, Kleine-Weber H, Schroeder S, et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell.* 2020;181(2):271-280. doi:10.1016/j.cell.2020.02.052

5. Watanabe Y, Allen JD, Wrapp D, McLellan JS, Crispin M. Site-specific glycan analysis of the SARS-CoV-2 spike. *Science.* 2020;369(6501):330-333. doi:10.1126/science.abb9983

6. Walls AC, Park Y-J, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell.* 2020;180:281-292.

7. Casalino L, Gaieb Z, Goldsmith JA, et al. Beyond shielding: the roles of glycans in the SARS-CoV-2 spike protein. *ACS Cent Sci.* 2020;6(10):1722-1734. doi:10.1021/acscentsci.0c01056

8. Cai Y, Zhang J, Xiao T, et al. Distinct conformational states of SARS-CoV-2 spike protein. *Science.* 2020;369(6511):1586-1592. doi:10.1126/science.abd4251

9. Yao H, Song Y, Chen Y, et al. Molecular architecture of the SARS-CoV-2 virus. *Cell.* 2020;183(3):730-738. doi:10.1016/j.cell.2020.09.018

10. Turoňová B, Sikora M, Schürmann C, et al. In situ structural analysis of SARS-CoV-2 spike reveals flexibility mediated by three hinges. *Science.* 2020;370(6513):203-208. doi:10.1126/science.abd5223

11. Choi YK, Cao Y, Frank M, et al. Structure, dynamics, receptor binding, and antibody binding of the fully glycosylated full-length SARS-CoV-2 spike protein in a viral membrane. *J Chem Theory Comput.* 2021;17(4):2479-2487. doi:10.1021/acs.jctc.0c01144

12. Deng Y, Liu J, Zheng Q, Yong W, Lu M. Structures and polymorphic interactions of two heptad-repeat regions of the SARS virus S2 protein. *Structure.* 2006;14(5):889-899. doi:10.1016/j.str.2006.03.007

13. Celigoy J, Ramirez B, Caffrey M. SARS-CoV heptad repeat 2 is a trimer of parallel helices. *Protein Sci.* 2011;20(12):2125-2129. doi:10.1002/pro.736

14. Muñoz V, Serrano L. Elucidating the folding problem of helical peptides using empirical parameters. *Nat Struct Mol Biol.* 1994;1(6):399-409. doi:10.1038/nsb0694-399

15. Vincent TL, Green PJ, Woolfson DN. LOGICOIL—multi-state prediction of coiled-coil oligomeric state. *Bioinformatics.* 2013;29(1):69-76. doi:10.1093/bioinformatics/bts648

16. Manalastas-Cantos K, Konarev PV, Hajizadeh NR, et al. ATSAS 3.0: expanded functionality and new tools for small-angle scattering data analysis. *J Appl Crystallogr.* 2021;54(1):343-355. doi:10.1107/S1600576720013412

17. Kabsch W. XDS. *Acta Crystallogr D Biol Crystallogr.* 2010;66(2):125-132. doi:10.1107/s0907444909047337

18. Sammito M, Millán C, Rodríguez DD, et al. Exploiting tertiary structure through local folds for crystallographic phasing. *Nat Methods.* 2013;10(11):1099-1101. doi:10.1038/nmeth.2644

19. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of *Coot. Acta Crystallogr Sect D Biol Crystallogr.* 2010;66(4):486-501. doi:10.1107/S0907444910007493

20. Adams PD, Afonine PV, Bunkóczi G, et al. *PHENIX* : a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr Sect D Biol Crystallogr.* 2010;66(2):213-221. doi:10.1107/S0907444909052925

21. Dürr E, Bosshard HR. Folding of a three-stranded coiled coil. *Protein Sci.* 2000;9(7):1410-1415. doi:10.1110/ps.9.7.1410

22. Wang X, Deng B, Sun Z. Thermodynamics of helix formation in small peptides of varying length in vacuo, in implicit solvent, and in explicit solvent. *J Mol Model.* 2019;25(1):3-undefined. doi:10.1007/s00894-018-3886-2

23. Loladze VV, Ermolenko DN, Makhatadze GI. Heat capacity changes upon burial of polar and nonpolar groups in proteins. *Protein Sci.* 2008;10(7):1343-1352. doi:10.1110/ps.370101

24. Liu J, Deng Y, Zheng Q, Cheng C-S, Kallenbach NR, Lu M. A parallel coiled-coil tetramer with offset helices [†,‡]. *Biochemistry.* 2006;45(51):15224-15231. doi:10.1021/bi061914m

25. Duquerroy S, Vigouroux A, Rottier PJM, Rey FA, Jan BB. Central ions and lateral asparagine/glutamine zippers stabilize the post-fusion hairpin conformation of the SARS coronavirus spike glycoprotein. *Virology.* 2005;335(2):276-285. doi:10.1016/j.virol.2005.02.022

26. Gurnon DG, Whitaker JA, Oakley MG. Design and characterization of a homodimeric antiparallel coiled coil. *J Am Chem Soc.* 2003;125(25):7518-7519. doi:10.1021/ja0357590

27. Gradišar H, Jerala R. *De novo* design of orthogonal peptide pairs forming parallel coiled-coil heterodimers. *J Pept Sci.* 2011;17(2):100-106. doi:10.1002/psc.1331

28. Ingallinella P, Bianchi E, Finotto M, et al. Structural characterization of the fusion-active complex of severe acute

respiratory syndrome (SARS) coronavirus. *Proc Natl Acad Sci.* 2004;101(23):8709-8714. doi:10.1073/pnas.0402753101

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.