Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

doi:10.1016/j.jmb.2010.07.043 J. Mol. Biol. (2010) 402, 619–628

Contents lists available at www.sciencedirect.com

Journal of Molecular Biology

journal homepage: http://ees.elsevier.com.jmb

COMMUNICATION

NMR Structure of the SARS-CoV Nonstructural Protein 7 in Solution at pH 6.5

Margaret A. Johnson[1], Kristaps Jaudzems[1] and Kurt Wüthrich[1,2,3]*

[1]*Department of Molecular Biology, The Scripps Research Institute, La Jolla, CA 92037, USA*
[2]*Department of Chemistry, The Scripps Research Institute, La Jolla, CA 92037, USA*
[3]*The Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, CA 92037, USA*

Received 11 June 2010;
received in revised form
22 July 2010;
accepted 23 July 2010
Available online
13 August 2010

Edited by M. F. Summers

Keywords:
severe acute respiratory syndrome;
coronavirus;
nsp7;
NMR structure;
conformational polymorphism

The NMR structure of the severe acute respiratory syndrome coronavirus nonstructural protein (nsp) 7 in aqueous solution at pH 6.5 was determined and compared with the results of previous structure determinations of nsp7 in solution at pH 7.5 and in the crystals of a hexadecameric nsp7/nsp8 complex obtained from a solution at pH 7.5. All three structures contain four helices as the only regular secondary structures, but there are differences in the lengths and sequence locations of the four helices, as well as between the tertiary folds. The present study includes data on conformational equilibria and intramolecular rate processes in nsp7 in solution at pH 6.5, which provide further insights into the polymorphisms implicated by a comparison of the three presently available nsp7 structures.

Introduction

The severe acute respiratory syndrome coronavirus (SARS-CoV) nonstructural protein (nsp) 7 is of interest for its potential roles in the transcription and replication of the positive-stranded viral RNA genome. The proteins nsp7–nsp10, which are conserved among all coronaviruses (CoVs) but have no functional homologs outside of *Coronaviridae*, are translated as part of the viral polyproteins pp1a and pp1ab, and mature proteins are released by the action of the SARS-CoV protease nsp5.[1–3] An important role of nsp7 is indicated by the observation that deletion of nsp7 or mutation of the nsp7/nsp8 proteolytic cleavage site is lethal to murine hepatitis virus (MHV).[4] The expression of nsp7 in infected cells has been demonstrated for viruses belonging to three of the major CoV phylogenetic groups, namely, human CoV 229E (group I),[5] MHV (group II),[6] and avian infectious bronchitis virus (group III)[7]. During infection, nsp7 localizes to membrane-related sites of viral replication in the cytoplasm;[5,6,8] and in MHV, it has been shown to interact specifically with nsp1 and nsp10 at sites of viral RNA synthesis.[8]

To provide a structural basis for functional studies, an NMR structure of nsp7 was determined in aqueous solution at pH 7.5,[9] and crystal structure

*Corresponding author. E-mail address: wuthrich@scripps.edu.

Abbreviations used: nsp, nonstructural protein; SARS-CoV, severe acute respiratory syndrome coronavirus; CoV, coronavirus; MHV, murine hepatitis virus; PDB, Protein Data Bank; NOE, nuclear Overhauser enhancement; Pf, protection factor; EGS, ethylene glycol bis[succinimidylsuccinate]; 3D, three-dimensional; NOESY, NOE spectroscopy.

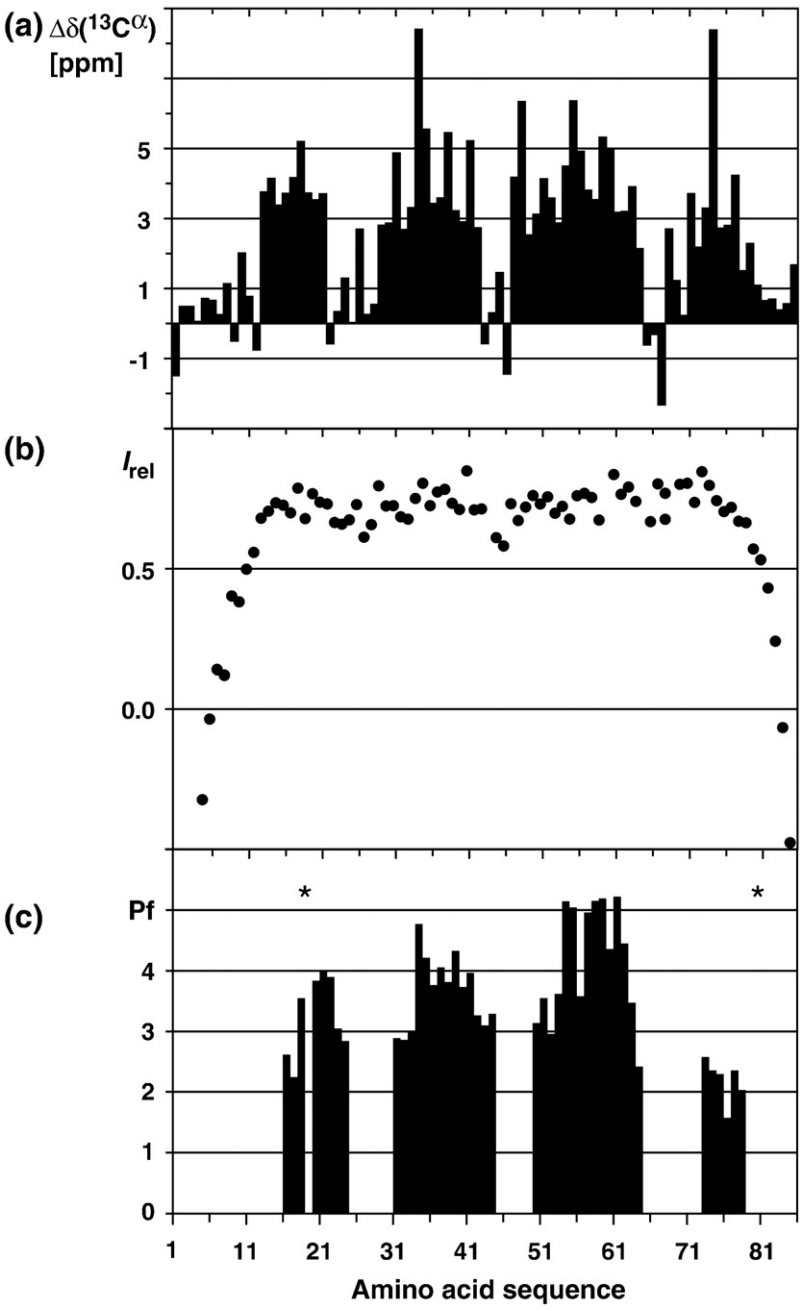

Fig. 1. Plots *versus* the nsp7 amino acid sequence of NMR data measured in aqueous solution at pH 6.5 and $T = 25$ °C. (a) $^{13}C^{\alpha}$ chemical shift deviations from random-coil values, $\Delta\delta(^{13}C^{\alpha})$. The sequence locations of the four α-helices are indicated at the top. (b) $^{15}N\{^{1}H\}$ NOE values (I_{rel}) measured on a Bruker Avance 600-MHz spectrometer with a TCI *z*-gradient cryoprobe using sensitivity-enhanced experiments[14,15] at a saturation period of 3.0 s and at a total interscan delay of 5.0 s. (c) Amide $^{1}H/^{2}H$ exchange protection factors. Asterisks indicate positions where no measurements could be made due to spectral overlap. Pf values were determined using a 2 mM ^{15}N-labeled protein sample that was lyophilized from $^{1}H_2O$ solution and then redissolved in 99.9% $^{2}H_2O$. The decay of the signal intensity of $^{15}N–^{1}H$ correlation peaks was monitored by acquiring a series of two-dimensional $^{13}N,^{1}H$ heteronuclear single quantum coherence spectra at different times after the preparation of the $^{2}H_2O$ solution. Each spectrum was acquired for 4–15 min, and the study was continued for 11 days. Peak intensities were fitted to an exponential equation of the form $I = I_0\exp(-k_{ex}t)$, where we accounted for residual peak intensities. Pf values were calculated by taking into account amino acid sequence effects on random-coil exchange rates (see the text).

determination was reported[10] for a complex with nsp8, which has been shown to have hexadecamer RNA primase activity.[11] The complex in the crystal contains eight molecules each of nsp7 and nsp8, forming a channel with appropriate width and charge to accommodate double-stranded RNA, and it was proposed to function as a potential processivity factor for RNA replication.[10] In both of these structures, the polypeptide fold of nsp7 includes four helices as the only regular secondary structures, which cover approximately 60% of the 85-residue polypeptide chain. A comparison of the nsp7 folds in the two structures revealed extensive differences in the sequence positions and lengths of the four helices, as well as in the spatial arrangement of the helices in the tertiary structure. Interestingly, both structure determinations led to the conclusion, based on Dali searches,[12,13] that nsp7 represented a novel fold.[9,10]

In view of the implicated conformational polymorphism and in consideration of the fact that the NMR structure at pH 7.5 had to be calculated from a scarce set of experimental conformational constraints,[9] we decided to further investigate the behavior of nsp7 in different environments. The present project was therefore initiated by the screening of a wide range of solution conditions

in search of protein samples that would enable the recording of high-quality NMR data and the collection of more extensive sets of conformational constraints than had been possible at pH 7.5 and high ionic strengths.[9] Based on the results of this screening, an aqueous solution at pH 6.5 and with low ionic strength was selected for a new NMR structure determination of nsp7. Additional NMR studies of conformational equilibria and dynamic processes further provided a foundation for rationalizing some aspects of variations among the three presently available nsp7 structures.

Preparation of an nsp7 solution at pH 6.5

Nsp7 was expressed in *Escherichia coli* BL21 (DE3)-RIL cells using the plasmid pET28a with an N-terminal 6×His tag and a tobacco etch virus protease cleavage site.[9] Two residues derived from the tag, Gly1 and His2, remained attached to the protein after purification. Uniformly ^{15}N-labeled or ^{13}C,^{15}N-labeled protein was produced by growth in M9 minimal medium containing 1 g/L $^{15}NH_4Cl$ as the sole nitrogen source and 4 g/L unlabeled glucose or $^{13}C_6$-D-glucose as the sole carbon source. Cell

Table 1. Input for the structure calculation and statistics of the ensemble of 20 energy-minimized CYANA conformers used to represent the NMR structure of nsp7 at pH 6.5 and T = 25 °C

Quantity[a]	Value[a]
NOE upper distance limits[b]	2035
Intraresidual	448
Short range	614
Medium range	599
Long range	374
Restraints/residue	28
Long-range restraints/residue	5
Dihedral angle constraints	382
Residual target function value (Å^2)	1.95±0.24
Residual NOE violations	
Number >0.1 Å	36±5
Maximum (Å)	0.15±0.01
Residual dihedral angle violations	
Number >2.5°	1±1
Maximum (°)	2.25±1.45
Amber energies (kcal/mol)	
Total	−3264.44±65.37
Van der Waals	−217.35±15.87
Electrostatic	−3725.29±57.16
RMSD from ideal geometry	
Bond lengths (Å)	0.0078±0.0002
Bond angles (°)	2.158±0.072
RMSD to the mean coordinates (Å)[c]	
bb	0.46±0.06 (11–82)
ha	0.82±0.08 (11–82)
Ramachandran plot statistics (%)[d]	
Most favored regions	75.2
Additionally allowed regions	20.7
Generously allowed regions	2.9
Disallowed regions	1.2

[a] The top eight entries describe the input from NMR experiments. The other entries refer to the ensemble of 20 CYANA conformers after energy minimization with OPALp. The ranges indicate standard deviations.

[b] Structure determination was based on a three-dimensional (3D) ^{15}N-resolved ^{1}H,^{1}H NOE spectroscopy (NOESY) spectrum with a 100 ms mixing time and on two 3D ^{13}C-resolved ^{1}H,^{1}H NOESY spectra with the carrier frequency centered in the aliphatic and aromatic carbon regions and with mixing times of 150 ms and 60 ms, respectively, recorded on a Bruker Avance 800 spectrometer with a TXI z-gradient probe. Protein backbone resonances were assigned based on 3D HNCA, 3D HNCACB, and 3D CBCA(CO)NH experiments.[16] Automated side-chain resonance assignments were based on the use of the three 3D NOESY data sets as input for the program ASCAN,[17] followed by interactive verification based on a 3D HC(C)H total correlation spectroscopy experiment. ^{1}H chemical shifts were referenced to internal 3-(trimethylsilyl)-1-propanesulfonic acid sodium salt (DSS). The ^{13}C and ^{15}N chemical shifts were referenced indirectly to DSS using absolute frequency ratios.[18] Structure calculation used the three aforementioned NOESY data sets as input for the stand-alone program suite ATNOS/CANDID 2.2[19,20] and the torsion angle molecular dynamics program CYANA 3.0.[21] Backbone φ and ψ dihedral angle constraints derived from $^{13}C^{\alpha}$ chemical shifts were used as supplementary data in the input.[22,23] In the seventh ATNOS/CANDID/CYANA cycle, 40 conformers were generated and subjected to energy minimization in a water shell with OPALp[24,25] using the AMBER force field,[26] and the 20 best energy-minimized conformers were selected to represent the solution structure. The program MOLMOL[27] was used for structure analysis and presentation. The stereochemical quality of the molecular models was analyzed using the PDB validation server (http://deposit.pdb.org/validate).

[c] bb indicates backbone N, C^{α}, and C′ atoms; ha stands for "all heavy atoms." The numbers in parentheses indicate the residues for which the RMSD was calculated.

[d] As determined by PROCHECK.[28]

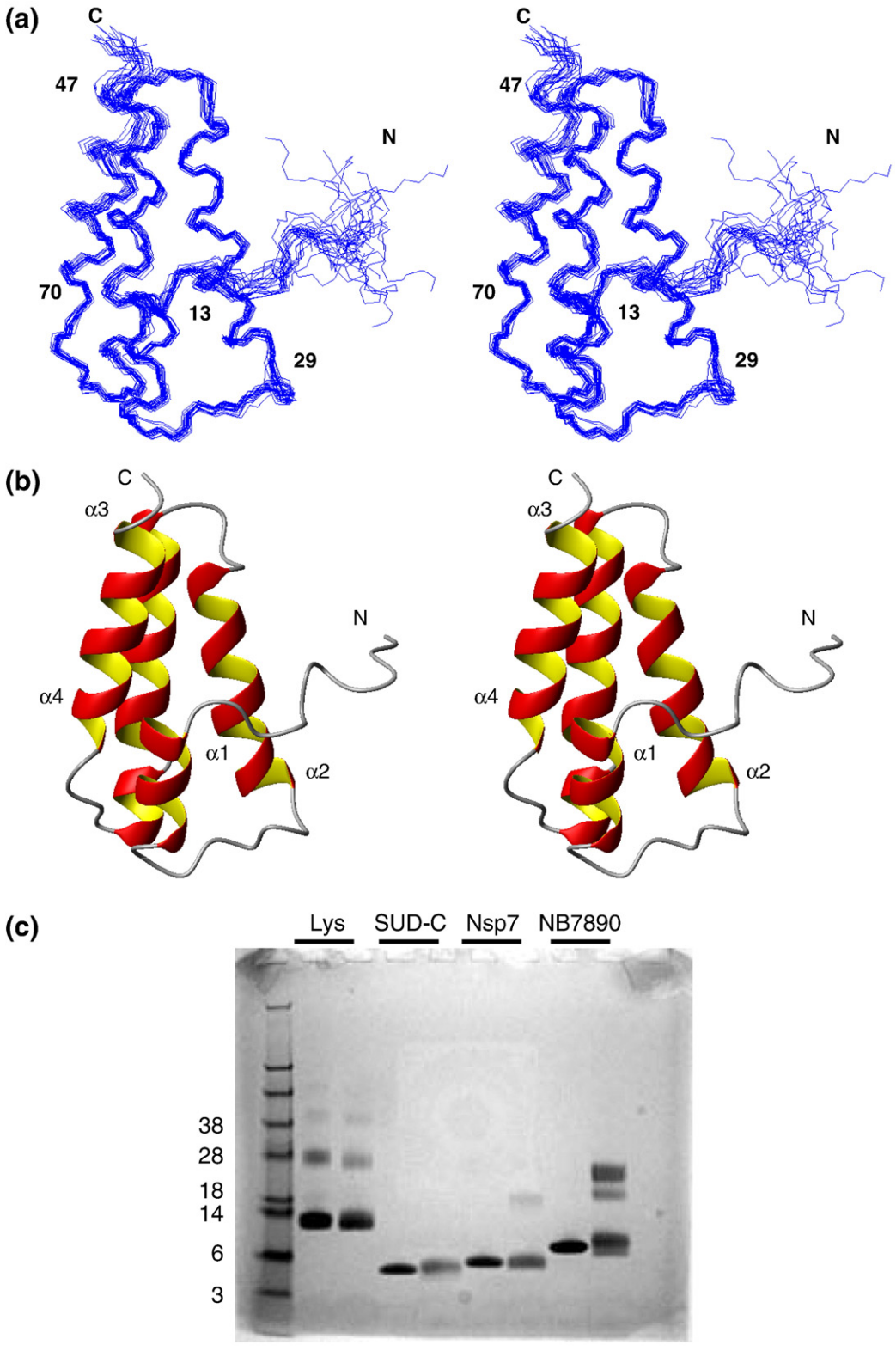

Fig. 2 (*legend on next page*)

cultures were grown at 37 °C, with shaking, to an optical density at 600 nm of ~0.6. The temperature was then lowered to 18 °C, expression was induced with 1 mM isopropyl-β-D-thiogalactopyranoside, and the cultures were grown for a further 18 h.

Following collection of the cells and storage at −80 °C, cell pellets were disrupted by sonication in lysis buffer [50 mM Tris (pH 8), 500 mM NaCl, 5 mM imidazole, 0.1% Triton X-100, 3.5 mM DTT, and ethylenediaminetetraacetic-acid-free Complete protease inhibitors (Roche)]. The solution was centrifuged to remove cell debris and filtered through a 0.22-μM syringe filter (Millipore), and the supernatant was applied to a 5-ml HisTrap Crude column (GE Life Sciences) equilibrated with buffer A (lysis buffer without detergent and protease inhibitors). The bound proteins were eluted with a gradient from 5 mM to 500 mM imidazole, concentrated, and exchanged by ultrafiltration (Millipore Ultrafree centrifugal concentrators; molecular weight cutoff, 3 000) into buffer A′ [50 mM sodium phosphate (pH 7.5) and 300 mM NaCl] containing 10 mM DTT. The solution was treated with tobacco etch virus protease for 18 h at room temperature, and the cleaved protein was then eluted in the flow-through of a 5-ml HisTrap Crude column equilibrated with buffer A′ containing 4 mM DTT (since nsp7 oxidizes readily, freshly prepared buffers and maximum DTT concentrations compatible with column resins were used throughout purification). The protein was again concentrated, filtered, and further purified on a Superdex 75 26/60 column (GE Life Sciences) equilibrated with buffer A′ containing 5 mM DTT. Finally, the protein was exchanged by ultrafiltration into 'NMR buffer' [50 mM sodium phosphate (pH 6.5) and 150 mM NaCl]. The NMR samples contained 2 mM nsp7, 10 mM DTT-d_{10}, 7% D_2O, and 0.02% NaN_3 in a volume of 600 μl.

Structure and dynamics of nsp7 in solution at pH 6.5

NMR samples containing 2 mM nsp7 in 50 mM sodium phosphate buffer (pH 6.5) with 150 mM NaCl were used for structure determination at 25 °C. This choice was based on the results of a screen for high-quality NMR spectra in a wide range of buffers, pH values, and salt concentrations, using ^{15}N-labeled nsp7 and 1.7-mm microcoil NMR equipment at 700 MHz, and on circular dichroism measurements of the unfolding temperature.

Complete backbone and side-chain assignments for nsp7—except for $^{15}N^{\delta 1}$, $H^{\delta 1}$, and $^{15}N^{\varepsilon 2}$ of the histidines, $^{15}N^{\varepsilon}$ and H^{ε} of the arginines, and all backbone $^{13}C'$ atoms—were obtained. $^{13}C^{\alpha}$ chemical shifts showed that nsp7 at pH 6.5 is an α-protein with four helices (Fig. 1a). Structure calculation then resulted in a high-quality NMR structure, as shown by the statistics for the final cycle of calculation (Table 1). The protein forms an antiparallel bundle of four helices (Fig. 2a and b). The polypeptide segment 1–10, which includes the tag-derived residues Gly1 and His2, forms a disordered tail leading to a short extended segment of residues 11–12, which packs against the side chains of helices α2 and α4. Helix α1 (residues 13–20) is linked by a well-defined loop of residues 21–28 with helix α2 (residues 29–42). A short segment of nonregular secondary structure leads to helix α3 (residues 47–65), and a short loop connects to helix α4 (residues 70–82). Multiple structural homologs of the structure of Fig. 2a and b were identified by a Dali search of the Protein Data Bank (PDB).[12,13] These structural similarities probably result from the broad distribution of the four-helix bundle as a folding motif among protein families, and it is unlikely that these structural homologs indicate previously undetected functional homologs of nsp7 outside of *Coronaviridae*. The protein was further characterized by chemical cross-linking (Fig. 2c), which showed that nsp7 is monomeric in solution, and by a steady-state $^{15}N\{^1H\}$ nuclear Overhauser enhancement (NOE) experiment, which is sensitive to the picosecond-to-nanosecond timescale mobility of the polypeptide backbone. Increased subnanosecond mobility was evident for the N-terminal decapeptide segment and the C-terminal hexapeptide, which includes the last turn of helix α4 (Fig. 1b), whereas the data for residues 13–79 show that this central polypeptide segment forms a compact globular fold.

Conformational equilibria in nsp7 solutions at pH 6.5

The amide $^1H/^2H$ exchange rates in ^2H_2O solutions of proteins that have been lyophilized from

Fig. 2. Solution structure of nsp7 at pH 6.5. (a) Stereo view of a bundle of 20 energy-minimized conformers superimposed for minimal RMSD of the backbone N, C^{α}, and C′ atoms of residues 11–82. The chain ends and the start positions of each α-helix are labeled. (b) Stereo ribbon drawing of the conformer with minimal RMSD to the mean coordinates of the ensemble in (a). The chain ends are identified, and α-helices are labeled at their N-termini. (c) SDS-PAGE gel showing the results of an ethylene glycol bis[succinimidylsuccinate] (EGS) cross-linking experiment. The left margin shows molecular mass standards (in kDa). The remaining four pairs of lanes show lysozyme (14.1 kDa), SUD-C (7.6 kDa), nsp7 (9.5 kDa), and NB7890 (11.6 kDa). Lysozyme and SUD-C are monomeric proteins, while NB7890 is a dimer. For each protein, the left lane shows the control solution with no EGS added, and the right lane shows the result of the cross-linking reaction. The reactions were carried out with 50 μ M protein solutions at pH 7.3 and 5 mM EGS for 75 s at room temperature, and stopped by addition of Tris buffer (pH 8.0) to a concentration of 53 mM.

^1H_2O reflect the degree of protection of each proton by protein secondary and tertiary structures. Protons that are involved in the hydrogen bonds of regular secondary structures or that are otherwise sequestered from the solvent experience lower rates of chemical exchange. Amide proton protection factor (Pf) is defined as $\log(k_{in}/k_{ex})$, where k_{ex} is the measured hydrogen/deuterium exchange rate constant, and k_{in} is the intrinsic $^1H/^2H$ exchange rate constant for the same residue type when exposed to the solvent.[29] In nsp7 at pH 6.5, helices α2 and α3 have the highest protection factors, followed by α1 and α4; for polypeptide segments with nonregular secondary structure, the protection was too small to be measured with the standard approach used here (Fig. 1c). The lack of protection for the amide protons of the N-terminal tetradecapeptide is in line with the presence of a short helix α1 in the NMR structure (Figs. 1 and 2).

Of special interest is the behavior of helix α4. The reduced $\Delta\delta(^{13}C^{\alpha})$ values of residues 78–82, compared to other helical regions (Fig. 1a), and the comparatively low protection factors for the entire helix (Fig. 1c) indicate that there is a reduced population of α4 in the NMR structure because of a dynamic equilibrium with unstructured solvent-accessible conformations. The latter are apparently not manifested in the NOE-based NMR structure due to the absence of short $^1H–^1H$ distances corresponding to d_{NN}, $d_{\alpha N}(i, i+3)$, and $d_{\alpha\beta}(i, i+3)$ in the helix.[30] Since subnanosecond timescale motion of the protein backbone was observed only for the C-terminal hexapeptide segment of the protein (Fig. 1b), there was an indication that the conformational equilibria involving helix α4 are governed by slower motions. This indication was confirmed by a line shape analysis. The residues Ser63 and Gln65 near the C-terminal end of α3, Val68 in the loop joining α3 and α4, and Asn71 and Leu73 at the start of α4 all exhibit a pronounced line broadening when compared with residues in molecular regions that are not directly affected by rate processes involving α4 (Fig. 3). This line broadening was more pronounced when the temperature was decreased from 308 K to 288 K, indicating that the exchange between α4 and conformations with solvent-exposed amide groups approaches the fast rate limit on the chemical shift timescale at 308 K (millisecond-to-submillisecond timescale) such that a single averaged signal is seen at all temperatures in Fig. 3.

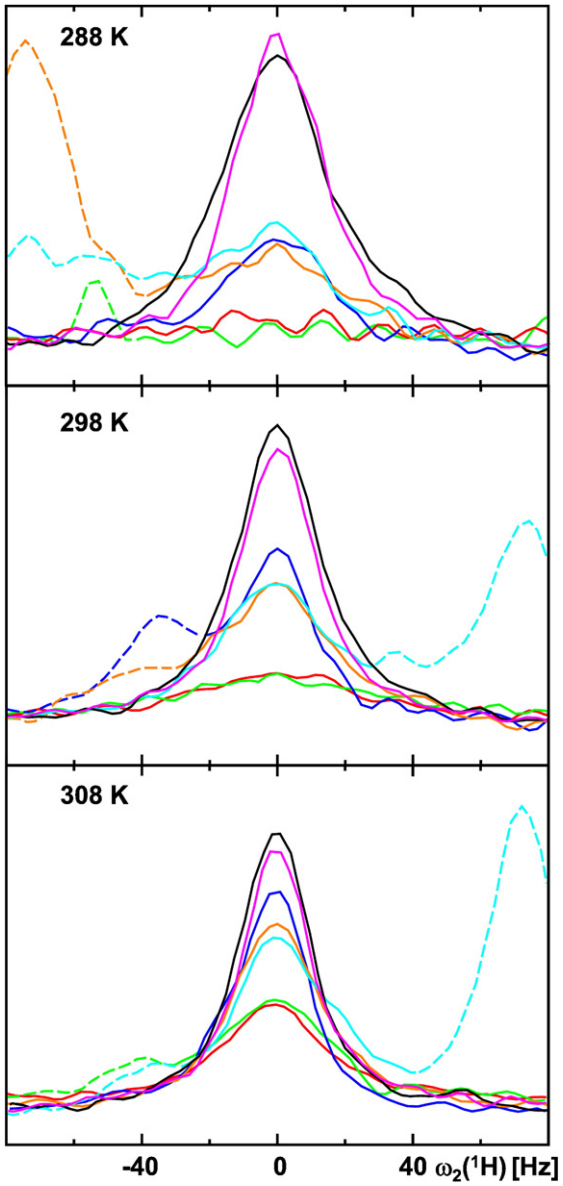

Fig. 3. Cross sections along ω_2 (^1H) from two-dimensional $^{15}N,^1H$ heteronuclear single quantum correlation spectra at 288 K, 298 K, and 308 K illustrating line broadening due to conformational exchange for the resonances of Ser63 (red), Gln65 (green), Val68 (blue), Asn71 (orange), and Leu73 (cyan). The signals of Ala32 (black) and Val60 (magenta) are shown as references without exchange line broadening. Those parts of the cross sections that are due to overlap with nearby peaks are drawn with broken lines. The digital resolution is 2.0 Hz/point.

Functional implications from the polymorphism of nsp7 structures in different environments

Overall, the present study confirms indications from earlier work that helices α2 and α3 form a conserved core of the nsp7 structure, with their lengths, positions, and relative orientations being largely preserved in different environments, and with helices α1 and α4 adopting quite different lengths, positions in the sequence, and relative orientations (Fig. 4). In addition, we obtained new information on the timescale of conformational equilibria in the solution structure at pH 6.5. A

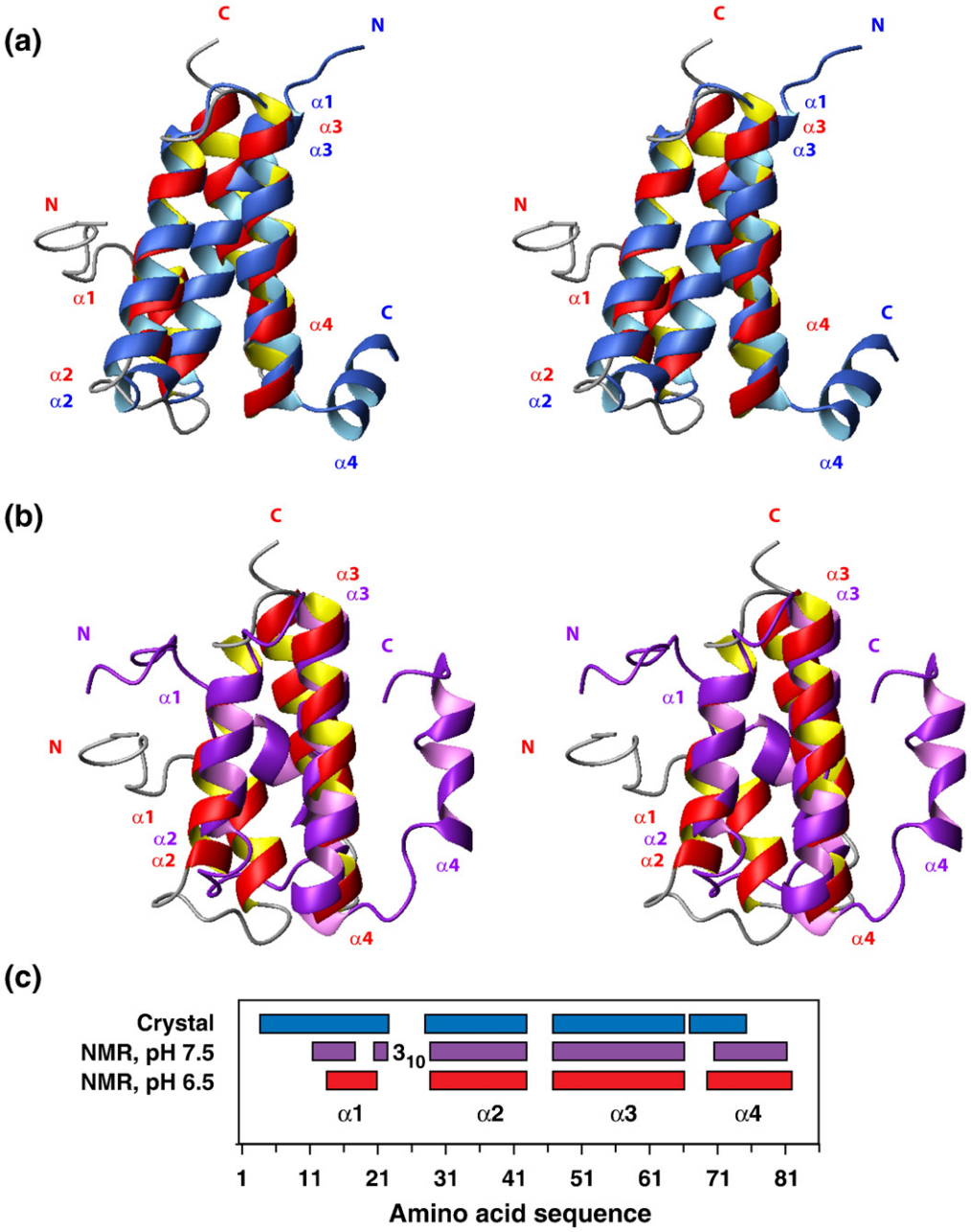

Fig. 4. Stereo ribbon drawings of the superpositions of the solution structure of nsp7 at pH 6.5 (red/yellow) with the two previously reported nsp7 structures, and sequence locations of α-helices in the three structures. (a) Superposition with the crystal structure of nsp7 in a complex with nsp8 (blue). (b) Superposition with the solution structure of nsp7 at pH 7.5 (purple). The structures were superimposed for minimal RMSD of the backbone N, C^{α}, and C′ atoms of helices α2 and α3 (residues 29–42 and 47–65), which yielded RMSD values of 1.63 Å (a) and 2.70 Å (b) [the corresponding values for the superposition of the C^{α} atoms of helices α2 and α3 are 1.77 Å (a) and 2.91 Å (b)]. The chain ends and the starting positions of the helices are labeled with the respective colors. (c) Locations of α-helices in the NMR structures determined at pH 6.5 and pH 7.5,[9] and in the crystal structure of the complex with nsp8.[10] The numbering of helices is indicated for the NMR structure at pH 6.5, and the color scheme is the same as in (a) and (b). The locations of helices in the pH 7.5 solution structure were taken from Peti *et al.*[9] In the pH 6.5 solution structure, the locations of helices were determined by an automatic analysis of the ensemble of 20 energy-minimized conformers (Fig. 2a) with the program MOLMOL,[27] which employs the algorithm of Kabsch and Sander[31] for secondary structure identification. Helix locations were assigned by determining the most common start points and end points of each helix in the ensemble of conformers. The locations of helices in the crystal structure were determined by an analysis of the coordinates (PDB accession code 2AHM[10]) with MOLMOL, using the second molecule in the asymmetric unit as the representative conformer.

closer look at the three presently available nsp7 structures reveals that the crystal structure of the complex with nsp8 differs from the pH 6.5 solution structure primarily by a rotation of helix α4 away from the α2/α3 core and by helical folding of the polypeptide segment of residues 3–12 (Fig. 4a). The long helix α1 in the crystal structure (Fig. 4c), which would not be compatible with the amide proton protection factors measured in solution at pH 6.5 (Fig. 1c), then occupies the position taken by α4 in the pH 6.5 solution structure (Fig. 4a). In the pH 7.5 solution structure, α4 is packed against α3 and has no contacts with α2, so that helices α2, α3, and α4 line up to form a flat three-helix sheet (Fig. 4b), with α1 packed at an angle of about 45° against α2 and α3.

The nsp7 constructs used for NMR and X-ray crystallographic studies included the expression-tag-derived N-terminal elongations GH and GPLGS, respectively.[9,10] These tag-derived residues do not form part of helix α1 in any of the structures (Fig. 4c), and the tag-derived residues did not all give rise to a clearly observable electron density in the crystal structure. These specific tag residues are nonetheless of interest, since in the position preceding an α-helix, both His and Ser have been observed to participate in helix-stabilizing N-capping interactions.[32,33] In two of the four nsp7 protomers in the crystal structure, hydrogen bonding between the side chain of the tag-derived Ser and helix α1, which might contribute to the stabilization of a long helix α1 in the crystals (Fig. 4c), is indeed possible. However, since the other two nsp7 protomers in the crystal do not display this interaction, and since the tag-derived residues are separated from α1 by regions of nonregular secondary structure in solution, the tag-derived residues do not appear to have a dominant role with regard to the variable regular secondary structures seen under different conditions (Fig. 4).

Viral proteins can be exposed to significant pH changes as they move between different cellular compartments during viral infection; therefore, pH-dependent structure variations under near-physiological conditions may be relevant to triggering different protein–protein interactions during viral replication. CoV replication is known to occur in double-membrane vesicles derived from the rough endoplasmic reticulum,[34] which has a pH of about 7.0,[35] or from the endoplasmic reticulum/Golgi intermediate compartment. In contrast, viral budding takes place not only in the endoplasmic reticulum/Golgi intermediate compartment but also in the Golgi apparatus,[36,37] which has a pH of about 6.5.[38] The regulation, transport, and assembly processes involved in the transition from genome replication to viral particle budding and particle maturation are not well understood,[34] and a pH-dependent conformational transition in one or more of the nsps (most of which have not been found in mature viral particles[39]) could be an essential step in viral genome packaging. That the antimalarial drug chloroquine has activity against SARS-CoV and other viruses, through mechanisms that include a pH increase in the normally acidic trans-Golgi network,[40,41] provides further indications for a role of pH in controlling viral infection.

In the case of nsp7, where neither of the two solution structures is compatible with the nsp8 binding mode observed in the crystal structure, it is tempting to speculate that nsp8 binds to a transient form of nsp7 in which helix α4 is unfolded, as implied by the data of Figs. 1c and 3. Binding to nsp8 would trigger the formation of the short helix α4 immediately after α3, as well as the formation of the long helix α1 (Fig. 4c), which then takes up the position occupied by α4 in the pH 6.5 solution structure (Fig. 4a). In the crystal structure, the locations of α1 and α4 are stabilized by numerous contacts with hydrophobic side chains of nsp8, in addition to intramolecular contacts within nsp7, and residues 79–85 are structurally disordered, as evidenced by the lack of electron density. The conformational variability of α4 was seen also in the complex with nsp8, since the different molecules in the asymmetric unit of the crystal structure have different orientations of helix α4.[10] There is thus an indication that locking of helix α4 of nsp7 in a conformation that prevents binding to nsp8 could impair the replication machinery of the virus, which might provide the lead for future drug designs.

Interestingly, while the nsp7 sequence in helices α1–α3 is highly conserved among all CoVs, the sequence of helix α4 is quite variable,[9] indicating that the nsp7 fold could be interpreted as a scaffold consisting of a three-helix bundle, with a fourth helix in equilibrium with transiently unfolded forms affecting the species specificity of physiological activity. Considering the transient unfolding of α4 at pH 6.5, the scarce interactions of α4 with the rest of the protein in the pH 7.5 solution structure, and the complete lack of such interactions in the crystal structure, we truncated nsp7 at the end of helix α3. Constructs consisting of residues 1–65, 1–66, 1–67, 1–69, 1–70, and 1–72 all yielded partly folded, poorly soluble proteins that were not amenable to structure determination by NMR (M. D. Geralt, P. Serrano, M.A.J., and K.W., unpublished data). Conformations of the type observed at pH 6.5 in which helix α4 associates intimately with the bulk of the protein thus seem to be essential for protein stability and folding, and hence for the functional integrity of nsp7.

Databank depositions

The chemical shifts of nsp7 at pH 6.5 were deposited in BioMagResBank† under accession

† http://www.bmrb.wisc.edu

number 16981. The atomic coordinates of the ensemble of 20 conformers representing the solution structure of nsp7 at pH 6.5 were deposited in the PDB‡ under accession code 2KYS.

Acknowledgements

We thank M. D. Geralt and Dr. P. Serrano for helpful discussions. This study was supported by National Institute of Allergy and Infectious Diseases/National Institutes of Health contract HHSN266200400058C ("Functional and Structural Proteomics of the SARS-CoV"; principal investigator, Dr. P. Kuhn), the Joint Center for Structural Genomics through National Institutes of Health/National Institute of General Medical Sciences grant U54-GM074898 (principal investigator Dr. I. A. Wilson), and a fellowship from the Latvian Institute of Organic Synthesis (K.J.). K.W. is the Cecil H. and Ida M. Green Professor of Structural Biology at The Scripps Research Institute.

References

1. Masters, P. S. (2006). The molecular biology of coronaviruses. *Adv. Virus Res.* **66**, 193–292.
2. Prentice, E., McAuliffe, J., Lu, X., Subbarao, K. & Denison, M. R. (2004). Identification and characterization of severe acute respiratory syndrome coronavirus replicase proteins. *J. Virol.* **78**, 9977–9986.
3. Gorbalenya, A. E., Enjuanes, L., Ziebuhr, J. & Snijder, E. J. (2006). Nidovirales: evolving the largest RNA virus genome. *Virus Res.* **117**, 17–37.
4. Deming, D. J., Graham, R. L., Denison, M. R. & Baric, R. S. (2007). Processing of open reading frame 1a replicase proteins nsp7 to nsp10 in murine hepatitis virus strain A59 replication. *J. Virol.* **81**, 10280–10291.
5. Ziebuhr, J. & Siddell, S. G. (1999). Processing of the human coronavirus 229E replicase polyproteins by the virus-encoded 3C-like proteinase: identification of proteolytic products and cleavage sites common to pp1a and pp1ab. *J. Virol.* **73**, 177–185.
6. Bost, A. G., Carnahan, R. H., Lu, X. T. & Denison, M. R. (2000). Four proteins processed from the replicase gene polyprotein of mouse hepatitis virus colocalize in the cell periphery and adjacent to sites of virion assembly. *J. Virol.* **74**, 3379–3387.
7. Ng, L. F., Xu, H. Y. & Liu, D. X. (2001). Further identification and characterization of products processed from the coronavirus avian infectious bronchitis virus (IBV) 1a polyprotein by the 3C-like proteinase. *Adv. Exp. Med. Biol.* **494**, 291–298.
8. Brockway, S. M., Lu, X. T., Peters, T. R., Dermody, T. S. & Denison, M. R. (2004). Intracellular localization and protein interactions of the gene 1 protein p28 during mouse hepatitis virus replication. *J. Virol.* **78**, 11551–11562.
9. Peti, W., Johnson, M. A., Herrmann, T., Neuman, B. W., Buchmeier, M. J., Nelson, M. *et al.* (2005). Structural genomics of the severe acute respiratory syndrome coronavirus: nuclear magnetic resonance structure of the protein nsp7. *J. Virol.* **79**, 12905–12913.
10. Zhai, Y., Sun, F., Li, X., Pang, H., Xu, X., Bartlam, M. & Rao, Z. (2005). Insights into SARS-CoV transcription and replication from the structure of the nsp7–nsp8 hexadecamer. *Nat. Struct. Mol. Biol.* **12**, 980–986.
11. Imbert, I., Guillemot, J. C., Bourhis, J. M., Bussetta, C., Coutard, B., Egloff, M. P. *et al.* (2006). A second, non-canonical RNA-dependent RNA polymerase in SARS coronavirus. *EMBO J.* **25**, 4933–4942.
12. Holm, L. & Sander, C. (1993). Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123–138.
13. Holm, L. & Sander, C. (1995). Dali: a network tool for protein structure comparison. *Trends Biochem. Sci.* **20**, 478–480.
14. Renner, C., Schleicher, M., Moroder, L. & Holak, T. A. (2002). Practical aspects of the 2D ^{15}N–{^{1}H}-NOE experiment. *J. Biomol. NMR*, **23**, 23–33.
15. Kay, L. E., Keifer, P. & Saarinen, T. (1992). Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J. Am. Chem. Soc.* **114**, 10663–10665.
16. Sattler, M., Schleucher, J. & Griesinger, C. (1999). Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. Nucl. Magn. Reson. Spectrosc.* **34**, 93–158.
17. Fiorito, F., Herrmann, T., Damberger, F. F. & Wüthrich, K. (2008). Automated amino acid side-chain NMR assignment of proteins using ^{13}C- and ^{15}N-resolved 3D [^{1}H,^{1}H]-NOESY. *J. Biomol. NMR*, **42**, 23–33.
18. Wishart, D. S., Bigam, C. G., Yao, J., Abildgaard, F., Dyson, H. J., Oldfield, E. *et al.* (1995). ^{1}H, ^{13}C and ^{15}N chemical shift referencing in biomolecular NMR. *J. Biomol. NMR*, **6**, 135–140.
19. Herrmann, T., Güntert, P. & Wüthrich, K. (2002). Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS. *J. Biomol. NMR*, **24**, 171–189.
20. Herrmann, T., Güntert, P. & Wüthrich, K. (2002). Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J. Mol. Biol.* **319**, 209–227.
21. Güntert, P., Mumenthaler, C. & Wüthrich, K. (1997). Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.* **273**, 283–298.
22. Spera, S. & Bax, A. (1991). Empirical correlation between protein backbone conformation and C^{α} and C^{β} ^{13}C nuclear magnetic resonance chemical shifts. *J. Am. Chem. Soc.* **113**, 5490–5492.
23. Luginbühl, P., Szyperski, T. & Wüthrich, K. (1995). Statistical basis for the use of $^{13}C^{\alpha}$ chemical shifts in protein structure determination. *J. Magn. Reson. B*, **109**, 229–233.

‡ http://www.rcsb.org/pdb

24. Koradi, R., Billeter, M. & Güntert, P. (2000). Point-centered domain decomposition for parallel molecular dynamics simulation. *Comput. Phys. Commun.* **124**, 139–147.
25. Luginbühl, P., Güntert, P., Billeter, M. & Wüthrich, K. (1996). The new program OPAL for molecular dynamics simulations and energy refinements of biological macromolecules. *J. Biomol. NMR*, **8**, 136–146.
26. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Jr, Ferguson, D. M. *et al.* (1995). A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **117**, 5179–5197.
27. Koradi, R., Billeter, M. & Wüthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graphics*, **14**, 51–55.
28. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291.
29. Bai, Y., Milne, J. S., Mayne, L. & Englander, S. W. (1993). Primary structure effects on peptide group hydrogen exchange. *Proteins*, **17**, 75–86.
30. Wüthrich, K. (1986). *NMR of Proteins and Nucleic Acids.* Wiley, New York, NY.
31. Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.
32. Davis, R. B., Jr & Lecomte, J. T. (2006). A dynamic N-capping motif in cytochrome *b*5: evidence for a pH-controlled conformational switch. *Proteins*, **63**, 336–348.
33. Aurora, R. & Rose, G. D. (1998). Helix capping. *Protein Sci.* **7**, 21–38.
34. Perlman, S. & Netland, J. (2009). Coronaviruses post-SARS: update on replication and pathogenesis. *Nat. Rev. Microbiol.* **7**, 439–450.
35. Tian, H., Klambt, D. & Jones, A. M. (1995). Auxin-binding protein 1 does not bind auxin within the endoplasmic reticulum despite this being the predominant subcellular location for this hormone receptor. *J. Biol. Chem.* **270**, 26962–26969.
36. Stertz, S., Reichelt, M., Spiegel, M., Kuri, T., Martinez-Sobrido, L., Garcia-Sastre, A. *et al.* (2007). The intracellular sites of early replication and budding of SARS-coronavirus. *Virology*, **361**, 304–315.
37. Salanueva, I. J., Carrascosa, J. L. & Risco, C. (1999). Structural maturation of the transmissible gastroenteritis coronavirus. *J. Virol.* **73**, 7952–7964.
38. Kim, J. H., Lingwood, C. A., Williams, D. B., Furuya, W., Manolson, M. F. & Grinstein, S. (1996). Dynamic measurement of the pH of the Golgi complex in living cells using retrograde transport of the verotoxin receptor. *J. Cell Biol.* **134**, 1387–1399.
39. Neuman, B. W., Joseph, J. S., Saikatendu, K. S., Serrano, P., Chatterjee, A., Johnson, M. A. *et al.* (2008). Proteomics analysis unravels the functional repertoire of coronavirus nonstructural protein 3. *J. Virol.* **82**, 5279–5294.
40. Keyaerts, E., Vijgen, L., Maes, P., Neyts, J. & Van Ranst, M. (2004). *In vitro* inhibition of severe acute respiratory syndrome coronavirus by chloroquine. *Biochem. Biophys. Res. Commun.* **323**, 264–268.
41. Vincent, M. J., Bergeron, E., Benjannet, S., Erickson, B. R., Rollin, P. E., Ksiazek, T. G. *et al.* (2005). Chloroquine is a potent inhibitor of SARS coronavirus infection and spread. *Virol. J.* **2**, 69.