Crucial steps in the structure determination of a coronavirus spike glycoprotein using cryo-electron microscopy

Alexandra Walls,† M. Alejandra Tortorici,‡ Berend-Jan Bosch,§ Brandon Frenz,† Peter J. M. Rottier,¶ Frank DiMaio,† Felix A. Rey,‡ and David Veesler†

†Department of Biochemistry, University of Washington, Seattle, Washington, 98195
‡Institut Pasteur, Département de Virologie, Unité de Virologie Structurale, Paris, France
§CNRS UMR 3569 Virologie, Paris, France
¶Virology Division, Department of Infectious Diseases and Immunology, Faculty of Veterinary Medicine, Utrecht University, 3584 CL Utrecht, The Netherlands

Received 31 August 2016; Accepted 22 September 2016
DOI: 10.1002/pro.3048
Published online 26 September 2016 proteinscience.org

Abstract: The tremendous pandemic potential of coronaviruses was demonstrated twice in the last 15 years by two global outbreaks of deadly pneumonia. Entry of coronaviruses into cells is mediated by the transmembrane spike glycoprotein S, which forms a trimer carrying receptor-binding and membrane fusion functions. Despite their biomedical importance, coronavirus S glycoproteins have proven difficult targets for structural characterization, precluding high-resolution studies of the biologically relevant trimer. Recent technological developments in single particle cryo-electron microscopy allowed us to determine the first structure of a coronavirus S glycoprotein trimer which provided a framework to understand the mechanisms of viral entry and suggested potential inhibition strategies for this family of viruses. Here, we describe the key factors that enabled this breakthrough.

Keywords: coronavirus spike protein; cryo-electron microscopy; rational vaccine design; rosetta; relion

Coronaviruses are enveloped viruses with large positive-sense RNA genomes. In humans, coronaviruses are responsible for up to 30% of respiratory tract infections including mild upper respiratory tract infections (common cold), croup, bronchiolitis and pneumonia.1 In addition, coronaviruses have fostered a lot of attention in the last 15 years due to the emergence of deadly viruses with tremendous pandemic potential: severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle-East respiratory syndrome coronavirus (MERS-CoV).1,2 After its first occurrence, SARS-CoV rapidly spread around the world, reaching all five continents and resulting in over 8096 cases and 774 deaths by July 2003.
The emergence of MERS-CoV in 2012 has resulted in the infection of 1800 people and 640 deaths as of today. Currently, there are no approved antiviral treatments or vaccines for any human coronavirus.

Coronaviruses use homotrimers of the spike (S) glycoprotein to promote cell attachment and fusion of the viral and host membranes. As it is virtually the only antigen present at the virus surface, S is the main target of neutralizing antibodies during infection and a focus of vaccine design. The coronavirus S is a class I viral fusion protein synthesized as a single chain precursor of ~1300 amino acids which trimerizes upon folding. It comprises an N-terminal S1 subunit containing the receptor-binding domain and a C-terminal S2 subunit which is the membrane-anchored stalk carrying out membrane fusion. Cleavage by furin-like host proteases at the junction between S1 and S2 (S2 cleavage site) occurs during biogenesis for some coronaviruses such as murine hepatitis virus (MHV, the prototypical and best studied coronavirus). Coronaviruses have proven difficult targets for structural characterization and all reported studies have provided atomic resolution data for only a few isolated domains. The SARS-CoV S has also been studied in its native environment by cryo-electron microscopy (cryoEM) of intact virions, providing insights at low resolution into its overall shape. However, the lack of high-resolution data for any coronavirus spike trimer until earlier this year had prevented a detailed analysis of the mechanisms associated with infection.

Single-particle cryoEM is an increasingly important technique in structural biology, which enables the study of biological macromolecules in a near-native environment. Cryo-EM is undergoing a technological revolution due to the development of direct detection cameras and dedicated algorithms for tracking beam-induced motion and stage drift in recorded movies. These advances led to an explosion of the number of high-resolution

Figure 1. 3D reconstruction of the MHV S trimer determined by single-particle cryoEM. A–B, 4.0 Å resolution 3D map colored by protomer. Two different views of the S trimer (from the side (A) and from the top, looking toward the viral membrane (B)) are shown. C–D, Ribbon diagrams showing the MHV S atomic model oriented and colored as in A–B.
structures determined using cryoEM worldwide for numerous proteins and protein complexes that had previously been intractable using other structural techniques.

We leveraged these recent advances to determine the first near-atomic resolution structure of a coronavirus S glycoprotein trimer earlier this year [Fig. 1(A–D)]. These results paved the way for understanding the mechanisms of infection of viruses responsible for outbreaks of deadly pneumonia such as SARS-CoV and MERS-CoV. This article provides an in-depth analysis of the key methodological aspects that made possible the determination of the structure of the MHV S ectodomain trimer. We attribute this success to three main factors which are the design of a pre-fusion stabilized construct, the strategy...
employed for cryoEM data collection and processing, and the availability of a recently developed de novo model building algorithm using Rosetta.\textsuperscript{23–25}

**Construct Design**

Viral fusion proteins adopt a metastable pre-fusion conformation at the virus surface until triggered to rearrange into a more stable post-fusion conformation which promotes merger of viral and host membranes.\textsuperscript{26} The significant magnitude of the conformational changes taking place during the fusion reaction could result in masking of epitopes initially accessible in the prefusion state and exposure of new epitopes specific to the post-fusion state. As a result, vaccine design initiatives aim at targeting the pre-fusion state of viral fusion proteins, which correspond to the conformation that could be detected by the immune system before infection.

![Figure 3. Micrographs of MHV S particles embedded in vitreous ice. (A) Particles showed signs of denaturation in regions of very thin ice likely due to excessive surface tension. (B) Data acquired in regions featuring slightly thicker ice than desired showed compact and well-folded particles, similar to what was observed using negative staining. Scale bar: 50 nm.](image1)

![Figure 4. Computational sorting of particle images using 3D classification. (A) CryoEM reconstructions corresponding to the four classes requested during unsupervised 3D classification using the Relion software without symmetry imposed. (B) Slices through the center of the 3D reconstructions shown in (A). Only the left two classes were retained for further processing.](image2)
The intrinsic metastability of viral fusion proteins is usually associated with challenges to preserve the pre-fusion state during purification. This is illustrated by the case of the respiratory syncytial virus (paramyxovirus) F protein which required co-expression of the ectodomain (fused to a T4 foldon motif) with a pre-fusion specific Fab to enable isolation of this conformation.27–29

During biogenesis, the MHV S protein is often naturally cleaved at the S1-S2 junction (S2 cleavage site) by Golgi-resident furin(-like) proteases3,30 [Fig. 2(A)] resulting in an increase in its fusogenic propensity. After cleavage, the S1 and S2 subunits remain non-covalently associated in the metastable pre-fusion S trimer. In the case of SARS-CoV and MERS-CoV, S2 processing has also been suggested to promote subsequent cleavage at a second site located just upstream of the fusion peptide (S2' cleavage site) to allow the fusion reaction to proceed upon virion uptake by a target cell.4,5 We engineered a construct featuring a single amino acid substitution in the S2 cleavage site to prevent furin processing and enhance the stability of the MHV S ectodomain pre-fusion structure. Substitution of an arginine residue present at position 717 by a serine residue at the site of cleavage (from RAHR to RAHS) resulted in the purification of a homogeneous uncleaved protein product as confirmed by SDS-PAGE analysis [Fig. 2(B)].

Although MHV S is known to oligomerize into homo-trimers upon translation in vivo, expression of the ectodomain yielded predominantly monomers, indicating that the transmembrane domain is required for trimerization and/or trimer stabilization. To promote oligomerization, an engineered trimerization motif based on the transcription factor GCN431,32 was C-terminally fused to the MHV S ectodomain in frame with the heptad repeat 2 (HR2) motif helix [Fig. 2(A)]. Biophysical analyses using analytical size exclusion chromatography coupled online to multi-angle light scattering33 (SEC-MALS) as well as native mass spectrometry confirmed the trimeric organization of the GCN4 stabilized MHV S ectodomain [Fig. 2(C)]. Proper folding of the purified MHV S ectodomain was confirmed by analyzing its binding affinity to the CEACAM1a ectodomain (the viral receptor) using microscale thermophoresis [Fig. 2(D)]. We determined a dissociation equilibrium constant of 48.5 ± 3.8 nM which is in good agreement with the value of 21.4 ± 4.2 nM reported by Peng et al.12 for the isolated receptor-binding domain. Imaging of this sample using negative staining EM further confirmed the homogeneity of the purified protein and suitability for high-resolution studies [Fig. 2(E)].

Cryo-EM Data Collection and Processing

Ice thickness has a strong influence on the final achievable resolution of single particle reconstructions. Ideally, the vitreous ice should be as thin as possible to still accommodate the particles of interest while maximizing Thon ring intensity at high spatial frequencies.34 Imaging was completed on a Titan Krios 300 kV microscope equipped with a Gatan K2 Summit direct electron detector operated in counting mode.18 Similarly to our previous work on the Thermoplasma acidophilum 20S proteasome, we initially
sought to acquire data from holes having the thinnest possible vitreous ice. However, the MHV S protein clearly showed signs of denaturation when images where acquired in such conditions [Fig. 3(A)]. We interpret this observation as resulting from the surface tension exerted on the S trimers in thin vitreous ice. Hence, we targeted holes with slightly thicker ice than desired in which we could observe compact well-folded MHV S trimers, similar to what was observed using negative staining EM [Fig. 3(B)]. We collected a large dataset (1,600 micrographs) at high defocus (2.0–5.0 μm) to maximize the low-resolution contrast and our ability to align the particle images during subsequent processing. This example illustrates that although it is not always possible to acquire data in the thinnest possible areas of ice, near-atomic resolution reconstructions can still be obtained by tailoring the imaging conditions appropriately.

One of the major challenges encountered during processing of cryoEM data is the presence of multiple 3D structures in a given dataset. These differences can result from different conformations of the same protein, different chemical compositions due to loss of one or several subunits of a protein complex, or (partial) denaturation of a fraction of the particles during purification or vitrification. If left untreated, this heterogeneity can limit the resolution and compromise the quality of the final map. 3D classification has emerged as an extraordinarily powerful tool to deal with structural heterogeneity in allowing to computationally isolate homogeneous subsets of the data.

We relied on extensive 2D and 3D classification using the Relion software to deal with the marked structural heterogeneity of the MHV S ectodomain trimer dataset. We ran a first round of 3D classification without imposing symmetry to improve separation of “good” and “compromised” particle images. Figure 4 shows isosurface representation [Fig. 4(A)] and slices [Fig. 4(B)] going through the center of each of the four reconstructions corresponding to the four classes requested during unsupervised 3D classification. Although looking at the

**Figure 6.** Hybrid modeling enabled atomic model building of the MHV S trimer. (A) An example of a disulfide bond rendered in green present in the S2 fusion machinery. The observation of numerous disulfide bonds resolved in the cryoEM reconstruction helped validate the register of the atomic model. (B) An example of a putative glycosylation site where additional density protrudes from the Asn 893 side chain. (C) Rosetta de novo placed a ~30 residue-long fragment that anchored the register of the model in the density for domains C and D which are characterized by weaker density than the central regions of the reconstruction. Bulky side chains are accounted for by the density and the map also shows additional density protruding from an asparagine residue corresponding to a putative glycosylation sequon. (D) An example of a putative glycosylation site where additional density protrudes from the Asn 657 side chain. Arrows indicate cryoEM density corresponding to putative glycans.
Fourier shell correlation (FSC 0.143) criterion [Fig. 5(A)]. The quality of the two maps, however, differed significantly as only the reconstruction computed after 3D classification showed features compatible with the resolution estimate [Fig. 5(B,C)]. This case study highlights that gold standard FSC measures internal consistency between two halves of the data, not resolution, and that the quality of the final map should always be in agreement with any numerical estimates of resolution. Starting from 1,200,000 particle images, we significantly reduced the size of the data set to 82,000 particles using 2D and 3D classification to generate the final 3D reconstruction at 4Å resolution showing well resolved α-helices, β-strands and amino acid side chains for a large part of the map [Fig. 5(B,C)].

Model Building

Obtaining an atomic model of the MHV S glycoprotein required a hybrid approach combining docking of available crystal structures, de novo modeling using Rosetta23,25,43,44 and Coot45,46 and density-guided homology modeling using RosettaCM.24

The C-terminal S2 subunit, which is the fusion machinery, is best defined in the density and was built using a combination of hand tracing with Coot and Rosetta de novo building.25 The observation of large, bulky side chain densities, several disulfide bonds resolved in the map and of density putatively corresponding to glycans for several asparagine residues belonging to N-glycosylation sequons were used as internal controls during model building [Fig. 6(A,B)]. The density corresponding to the N-terminal receptor-binding S1 subunit is not as well resolved as for the fusion machinery and features various levels of resolution in the reconstruction. The availability of two crystal structures for domain A12,47 (including a structure of the MHV domain A) and of several crystal structures for domain B10,11 was of tremendous assistance and allowed us to directly dock these models into the reconstruction. RosettaCM was then used to rebuild the core β-sheet of domain B and to derive a putative model (using density-guided homology modeling) for the disordered extension corresponding to the receptor-binding motifs in MERS-CoV and SARS-CoV. The quality of the map corresponding to domains C and D hampered manual sequence assignment for this region of the protein. Rosetta de novo25 successfully identified a ~30 residue-long fragment allowing to anchor the sequence register for domains C and D. The placement of several bulky side chains accounted for by the density and the identification of putative N-linked glycans suggested correct assignment, and allowed completion of the model [Fig. 6(C,D)]. The density for the linker connecting the S1 and S2 subunits is poorly resolved and Rosetta de novo was used to generate a putative model of this region of the protein which should be analyzed cautiously, as suggested by the high B-factors associated with it.

Discussion

In addition to recent developments in direct detector technology, the determination of the first near-atomic resolution structure of a coronavirus spike glycoprotein trimer was made possible by (i) engineering a pre-fusion stabilized ectodomain construct, (ii) using extensive computational classification of particle images to sort out sample heterogeneity and (iii) relying on major advances in the Rosetta automated model building algorithm. To conclude, our results allowed the identification of a conserved neutralizing epitope at the surface of the protein and suggested potential vaccinology strategies to elicit broadly neutralizing antibodies against coronaviruses.22 This could pave the way toward the development of the first vaccine against human coronaviruses.

References

1. Coleman CM, Frieman MB (2014) Coronaviruses: important emerging human pathogens. J Virol 88: 5209–5212.
2. Du L, He Y, Zhou Y, Liu S, Zheng BJ, Jiang S (2009) The spike protein of sars-cov—a target for vaccine and therapeutic development. Nat Rev Microbiol 7: 226–236.
3. Bosch BJ, van der Zee R, de Haan CA, Rottier PJ (2003) The coronavirus spike protein is a class i virus fusion protein: structural and functional characterization of the fusion core complex. J Virol 77:8801–8811.
4. Millet JK, Whittaker GR (2014) Host cell entry of middle east respiratory syndrome coronavirus after two-step, furin-mediated activation of the spike protein. Proc Natl Acad Sci USA 111:15214–15219.
5. Millet JK, Whittaker GR (2015) Host cell proteases: critical determinants of coronavirus tropism and pathogenesis. Virus Res 202:120–134.
6. Xu Y, Liu Y, Lou Z, Qin L, Li X, Bai Z, Pang H, Tien P, Gao GF, Rao Z (2004) Structural basis for coronavirus-mediated membrane fusion. Crystal structure of mouse
hepatitis virus spike protein fusion core. J Biol Chem 279:30514–30522.
7. Duquerroy S, Vigouroux A, Rottier PJ, Rey FA, Bosch BJ (2005) Central ions and lateral asparagine/glutamine zippers stabilize the post-fusion hairpin conformation of the sars coronavirus spike glycoprotein. Virology 345:276–285.
8. Gao J, Lu G, Qi J, Li Y, Wu Y, Deng Y, Geng H, Li H, Wang Q, Xiao H, Tan W, Yan J, Gao GF (2013) Structure of the fusion core and inhibition of fusion by a heptad repeat peptide derived from the s protein of middle east respiratory syndrome coronavirus. J Virol 87:13134–13140.
9. Supekar VM, Bruckmann C, Ingallinella P, Bianchi E, Pessi A, Carfi A (2004) Structure of a proteolytically resistant core from the severe acute respiratory syndrome coronavirus s2 fusion protein. Proc Natl Acad Sci USA 101:17958–17963.
10. Lu G, Hu Y, Wang Q, Qi J, Gao F, Li Y, Zhang Y, Zhang W, Yuan Y, Bao J, Zhang B, Shi Y, Yan J, Gao GF (2013) Molecular basis of binding between novel human coronavirus mers-cov and its receptor cd26. Nature 500:227–231.
11. Li F, Li W, Farzan M, Harrison SC (2005) Structure of sars coronavirus spike receptor-binding domain complexed with receptor. Science 309:1864–1868.
12. Peng G, Sun D, Rajeshankar KR, Qian Z, Holmes KV, Li F (2011) Crystal structure of mouse coronavirus receptor-binding domain complexed with its murine receptor. Proc Natl Acad Sci USA 108:10696–10701.
13. Wu K, Li W, Peng G, Li F (2009) Crystal structure of nls3 respiratory coronavirus receptor-binding domain complexed with its human receptor. Proc Natl Acad Sci USA 106:19970–19974.
14. Beniac DR, Andonov A, Grudeski E, Booth TF (2006) Architecture of the sars coronavirus prefusion spike. Nat Struct Mol Biol 13:751–752.
15. Beniac DR, deVarennes SL, Andonov A, He R, Booth TF (2007) Conformational reorganization of the sars coronavirus spike following receptor binding: implications for membrane fusion. PLoS One 2:e1082.
16. Campbell MG, Cheng A, Brilot AF, Moeller A, Lyumkins D, Veesler D, Pan J, Harrison SC, Potter CS, Carragher B, Grigorieff N (2012) Movies of ice-embedded particles enhance resolution in electron cryo-microscopy. Structure 20:1823–1828.
17. Veesler D, Campbell MG, Cheng A, Fu CY, Murez Z, Johnson JE, Potter CS, Carragher B (2013) Maximizing the potential of electron cryomicroscopy data collected using direct detectors. J Struct Biol 184:193–202.
18. Li X, Mooney P, Zheng S, Booth CR, Braunfeld MB, Gubbens S, Agard DA, Cheng Y (2013) Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-em. Nat Methods 10:584–590.
19. Bai XC, Fernandez IS, McMullan G, Scheres SH (2013) Ribosome structures to near-atomic resolution from thirty thousand cryo-em particles. Elife 2:e00461.
20. Grant T, Grigorieff N (2015) Measuring the optimal exposure for single particle cryo-em using a 2.6 a reconstruction of rotavirus vp6. Elife 4:e06980.
21. Scheres SH (2014) Beam-induced motion correction for sub-megadalton cryo-em particles. Elife 3:e03665.
22. Walls AC, Tortorici MA, Bosch BJ, Frenz B, Rottier PJ, DiMaio F, Rey FA, Veesler D (2016) Cryo-electron microscopy structure of a coronavirus spike glycoprotein trimmer. Nature 531:114–117.
23. DiMaio F, Song Y, Li X, Brunner MJ, Xu C, Conticello V, Egelman E, Marlovits TC, Cheng Y, Baker D (2015) Atomic-accuracy models from 4.5-Å cryo-electron microscopy data with density-guided iterative local refinement. Nat Methods 12:361–365.
24. Song Y, DiMaio F, Wang Y, Kim D, Miles C, Brunette T, Johnson J, Thompson J, Baker D (2013) High-resolution comparative modeling with rossetacman. Structure 21:1735–1742.
25. Wang Y, Kudryashev M, Li X, Egelman EH, Basler M, Cheng Y, Baker D, DiMaio F (2015) De novo protein structure determination from near-atomic-resolution cryo-em maps. Nat Methods 12:335–338.
26. Harrison SC (2008) Viral membrane fusion. Nat Struct Mol Biol 15:690–698.
27. McLeans JS, Chen M, Leung S, Graepel KW, Du X, Yang Y, Zhou T, Baxa U, Yasuda E, Beaumont T, Kumar A, Modjarrad K, Zheng Z, Zhao M, Xin N, Kwong PD, Graham BS (2013) Structure of rsv fusion glycoprotein trimmer bound to a prefusion-specific neutralizing antibody. Science 340:1113–1117.
28. McLeans JS, Yang Y, Graham BS, Kwong PD (2011) Structure of respiratory syncytial virus fusion glycoprotein in the postfusion conformation reveals preservation of neutralizing epitopes. J Virol 85:7788–7796.
29. Swanson KA, Settembre EC, Shaw CA, Dey AK, Rappuoli R, Mandl CW, Dormitzer PR, Carfi A (2011) Structural basis for immunization with postfusion respiratory syncytial virus fusion f glycoprotein (rsv f) to elicit high neutralizing antibody titers. Proc Natl Acad Sci USA 108:9619–9624.
30. Burkard C, Verheije MH, Wicht O, van Kasteren SI, van Kuppeveld FJ, Haagmans BL, Pelkmans L, Rottier PJ, Bosch BJ, de Haan CA (2014) Coronavirus cell entry occurs through the endo-/lysosomal pathway in a proteolysis-dependent manner. PLoS Pathog 10:e1004502.
31. Yin HS, Wen X, Paterson RG, Lamb RA, Jardetzky TS (2006) Structure of the parainfluenza virus 5 f protein in its metastable, prefusion conformation. Nature 439:38–44.
32. Eckert DM, Malashkevich VN, Kim PS (1998) Crystal structure of gen4-piq, a trimeric coiled coil with buried polar residues. J Mol Biol 284:859–865.
33. Veesler D, Blangy S, Siponen M, Vincentelli R, Cambillau C, Sciarra G (2009) Production and biophysical characterization of the cora transporter from methanospirillum hungatii enansarcina mazei. Anal Biochem 388:115–121.
34. Wu S, Armache JP, Cheng Y (2016) Single-particle cryo-em data acquisition by using direct electron detection camera. Microscopy 65:35–41.
35. Campbell MG, Veesler D, Cheng A, Potter CS, Carragher B (2015) 2.8 Å resolution reconstruction of the thermoplasma acidophilum 20s proteasome using cryo-electron microscopy. Elife 4:e06380.
36. Lyumkins D, Brilot AF, Theobald DL, Grigorieff N (2013) Likelihood-based classification of cryo-em images using freelign. J Struct Biol 183:377–388.
37. Scheres SH, Gao H, Valle M, Herman GT, Eggermont PP, Frank J, Carazo JM (2007) Disentangling conformational states of macromolecules in 3d-em through likelihood optimization. Nat Methods 4:27–29.
38. Bai XC, Rajendra E, Yang G, Shi Y, Scheres SH (2015) Sampling the conformational space of the catalytic subunit of human gamma-secretase. Elife 4:e11182.
39. Scheres SH (2012) Relion: implementation of a Bayesian approach to cryo-em structure determination. J Struct Biol 180:519–530.
40. Scheres SH (2012) A Bayesian view on cryo-em structure determination. J Mol Biol 415:406–418.
41. Egelman EH (2014) Ambiguities in helical reconstruction. Elife 3:e04969.
42. Subramaniam S, Earl LA, Falconieri V, Milne JL, Egelman EH (2016) Resolution advances in cryo-em enable application to drug discovery. Curr Opin Struct Biol 41:194–202.
43. DiMaio F, Leaver-Fay A, Bradley P, Baker D, Andre I (2011) Modeling symmetric macromolecular structures in rosetta3. PLoS One 6:e20450.
44. DiMaio F, Zhang J, Chiu W, Baker D (2013) Cryo-em model validation using independent map reconstructions. Protein Sci 22:865–868.
45. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of coot. Acta Cryst D66:486–501.
46. Brown A, Long F, Nicholls RA, Toots J, Emsley P, Murshudov G (2015) Tools for macromolecular model building and refinement into electron cryo-microscopy reconstructions. Acta Cryst D71:136–153.
47. Peng G, Xu L, Lin YL, Chen L, Pasquarella JR, Holmes KV, Li F (2012) Crystal structure of bovine coronavirus spike protein lectin domain. J Biol Chem 287:41931–41938.