nature research

Peer Review File

Manuscript Title: Structure of replicating SARS-CoV-2 polymerase

Reviewer Comments & Author Rebuttals

Reviewer Reports on the Initial Version:

Referee #1:

The authors report a novel structure of the SARS-CoV2 RdRp complex made of nsp7 and two copies of nsp8, using Cryo-EM at 2.9 A resolution.
They observe a new arrangement of the complex, which explains the unusual processivity of the RdRp engaged into RNA synthesis of the long Coronavirus genomes.

The work is original, and will be a landmark in SARS-CoV structural biology.

The authors did an excellent job to obtain a stable form of the complex.
The cryo-EM experiment take into account a huge raw data set (1.3 million particles), and the subsequent rigorous analysis of these data give rise to a very large number of (418000) particles corresponding to the nsp12-nsp8-nsp7-RNA complex.

The good quality of the samples as well the great number of effective particles measured allows to significantly limit experimental bias, with appropriate statistical methods, randomization procedures, as well as blind allocation strategies.
The reconstruction of the differents parts of the complexe are done rigorously.
As a result, the EM structures presented in this article are of very high quality.

The form of the observed complex is of great importance in the understanding of this highly processive RNA synthesis machine. The discovery of the nsp8 as sliding poles is unexpected, and makes great sense.

The authors provide appropriate credit of previous work by others.

minor point:
line 60: remove de novo in the sentence "...a short RNA during de novo synthesis". Indeed, there are no indications nor data that de novo synthesis occurs, if what is meant is "de novo initiation".

Bruno Canard

Referee #2 :

In their work, Hillen et al. present a structure of the replication complex of SARS-CoV-2 including extensive RNA and additional helical extensions of nsp8. While the structural work is generally of high quality, the model appears to be overbuilt, containing many modeled regions which are hard to justify in the density with confidence. The density of RNA and the polymerase active site is impressive and certainly of higher quality than other available structures. The text itself is incredibly brief and aside from the improved protein and RNA observations, is lacking major conclusions or a real discussion of the results.

Figure 1c: the gel needs nucleotide size markers

Line 43: reword to indicate that different RNAs were used for the primer extension and the structural work

Fig 3c: There is no density to support any of the indicated interactions for chain B

Line 233: unlikely that E. coli were grown to an OD of 600, correct typo
Line 261 and 285: indicate amount of TEV protease used in % w/w, also indicate temperature of digestion

Line 68: authors should be clearer in their reference to flexible nsp8 regions as they used one of the cited papers to dock into their EM density.

Line 72: indicate upon first mention which protein K58 is in and specifically what mutation is lethal.

Line 310: indicate length of gel

Line 325: indicate final protein concentration prior to freezing

Impressive RNA density

Model: there is no confidence in the sequence assignment or build for 2-30 of chain B. In the composite map used for refinement there is almost no density until residue 74 and this region of chain B only become visible upon low pass filtering. Similarly for chain D, there is no confident side chain density before residue 40. The same can be said for the N-terminal region of chain A. There is also a known 9 aa register shift in the C-terminus of nsp12 (Tristan Croll describes it on Twitter, @CrollTristan) that the authors should fix in their model.

The supplementary movies do not add to the data or conclusions already presented.

Referee #3:

The short manuscript by Hillen et al describes a single particle cryoEM analysis of SARS-CoV-2 nsp12-nsp7-nsp8 polymerase complex in complex with dsRNA that mimics a replication complex. The structure reported shows a 1:1:2 complex of nsp12-nsp7-nsp8 with the core at an average resolution of 2.9 Å. Focused refinement reveals the structure (at an average resolution of 3.5 Å) of the long helical N-terminal extensions of each of the nsp8 molecules contacting a pseudo-continuous dsRNA extending ~30 base-pairs away from the polymerase active site. The occurrence of this long RNA duplex appears to be fortuitous, resulting from the concatenation of duplexes of the input template whose sequence is partially palindromic. No NTPs were added to the sample so there has been no active RNA synthesis and the +1 site for the incoming nucleotide is vacant.
There are two recently published cryoEM structures of the SARS-CoV-2 nsp12-nsp7-nsp8 polymerase complex, the apo-structure by Gao et al (Science 2020) and the product-template complex with incorporated remdesivir monophosphate (nominally at 2.5 Å resolution) by Yin et al (bioRxiv at the time of writing of the manuscript, now in Science since 1 May 2020).
The added value of the Hillen et al work is the first visualisation, in the context of the polymerase complex, of the helical extensions of both nsp8 molecules (these extensions are disordered in the other polymerase complex structures) which run parallel to the outgoing RNA duplex making numerous backbone contacts via mainly basic residues to both strands. Amongst these contacting residues are nsp8 K58 whose mutation to alanine has previously been shown (Subissi et al, 2014) for the highly homologous SARS-CoV-1 system to seriously affect the processivity of the polymerase. This supports the interpretation of Hillen et al that the helical extensions act as electrostatic guide through which the product template RNA duplex is extruded and that help the polymerase to grip the RNA beyond which is possible for the core nsp12 polymerase thus promoting processivity.

The paper is clearly written although some points need more discussion (see below). The methods are particularly well explained. The figures and movies are good.
In the light of the current high interest in fully understanding the mechanism of the SARS-CoV-2 polymerase and the fact that this work was done independently and in parallel with the other recently published work, this novel insight makes the paper worth publishing in Nature. However, the following points need to be taken into account.
(1) The title is misleading. This is not the structure of an actively replicating polymerase since no RNA synthesis has occurred. It is more correct to say it is mimicking an actively replicating polymerase
(2) To emphasise point (1), the authors need to explain in the main text the nature and origin of the dsRNA in their structure and include Ext. Data Figure 1c in the main Figure 2 or 3. As it stands, Fig 3a is misleading as it suggests the RNA is continuous which does not square with the breaks in the RNA backbone visible in the structure figures.
(3) The authors should include a figure superposing nsp8a and nsp8b showing the asymmetry.
(4) The authors show multiple basic residues from the nsp8 helices interacting with the RNA. How can they explain that mutating only K58 has such a big effect.
(5) Subissi et al, 2014 showed that the exonuclease-methyltransferase nsp14 protein binds to the nsp12-nsp7-nsp8 polymerase complex (for SARS-CoV-1). The authors should mention this and comment on how proof reading might occur if during processive elongation, there is such a long double stranded region.
(6) Do the authors believe that product and template form a a complete dsRNA over their whole length or is there a mechanism to separate the strands at some point. The authors should discuss how their observation fits into the overall mechsim of genome replication.
(7) The authors use CTF refinement to improve the final cryo-EM density. Could they further improve the density by microscope optical aberration corrections implemented in RELION 3.1?
(8) The authors use unsupervised 3D classification to identify more defined 'sliding poles' class. Could focus classification (perhaps without alignment) on the same region reveal an even better defined class?
(9) Line 356 claims extraction at 1.3 Å/px, the final table 0.834 Å/px
(10)

Author Rebuttals to Initial Comments:

Referee #1 (Remarks to the Author):

The authors report a novel structure of the SARS-CoV2 RdRp complex made of nsp7 and two copies of nsp8, using Cryo-EM at 2.9 A resolution.

They observe a new arrangement of the complex, which explains the unusual processivity of the RdRp engaged into RNA synthesis of the long Coronavirus genomes.

The work is original, and will be a landmark in SARS-CoV structural biology.

The authors did an excellent job to obtain a stable form of the complex.

The cryo-EM experiment take into account a huge raw data set (1.3 million particles), and the subsequent rigorous analysis of these data give rise to a very large number of (418000) particles corresponding to the nsp12-nsp8-nsp7-RNA complex.

The good quality of the samples as well the great number of effective particles measured allows to significantly limit experimental bias, with appropriate statistical methods, randomization procedures, as well as blind allocation strategies.

The reconstruction of the differents parts of the complexe are done rigorously.

As a result, the EM structures presented in this article are of very high quality.

The form of the observed complex is of great importance in the understanding of this highly processive RNA synthesis machine. The discovery of the nsp8 as sliding poles is unexpected, and makes great sense.

The authors provide appropriate credit of previous work by others.

We thank the reviewer for the careful reading, appreciation and support.

minor point:

line 60: remove de novo in the sentence "...a short RNA during de novo synthesis". Indeed, there are no indications nor data that de novo synthesis occurs, if what is meant is "de novo initiation".

We agree and have deleted 'de novo'.

Referee #2 (Remarks to the Author):

In their work, Hillen et al. present a structure of the replication complex of SARS-CoV-2 including extensive RNA and additional helical extensions of nsp8. While the structural work is generally of high quality, the model appears to be overbuilt, containing many modeled regions which are hard to justify in the density with confidence. The density of RNA and the polymerase active site is impressive and certainly of higher quality than other available structures. The text itself is incredibly brief and aside from the improved protein and RNA observations, is lacking major conclusions or a real discussion of the results.

We thank the reviewer for the careful analysis of our data and the helpful comments. Note that the few protein regions in question could be modelled accurately due to the availability of prior high-resolution structural information for nsp8 regions. We edited the corresponding description in methods to make this clearer. See also our answers below. With respect to additional discussion, we wish to be careful and not to add more discussion at this time.

Figure 1c: the gel needs nucleotide size markers

Since this is separation of fluorescently labeled RNA, markers are not readily available. However, they are not needed in this particular case to draw the conclusion that the RdRp enzyme is active, because

only the original RNA (lower band) and the extended RNA resulting from RNA synthesis (upper band) are compared. Thus there is no ambiguity.

Line 43: reword to indicate that different RNAs were used for the primer extension and the structural work

Compare comment 2 by reviewer 3. We now show the nature of the template for structural studies in Figure 3, and also explain this better in the main text and in the figure legends.

Fig 3c: There is no density to support any of the indicated interactions for chain B

Although the density in this region clearly shows the secondary structure, it generally does not reveal side chains. We therefore did not indicate interactions in Fig. 3c as would typically be done with dashed lines, and we had noted in the legend that we show here residues in 'proximity' to RNA rather than residues 'contacting' RNA. However, as the reviewer is obviously concerned that the reader may mistake our figure as evidence for such detailed knowledge, we also changed Fig. 3a to avoid such misunderstanding. To account for the remaining uncertainty of the side chain-RNA nucleotide interactions, we now indicate in the figure panel 3a those RNA nucleotides that could be contacted by the corresponding residues for which side chain density is lacking (horizontal lines). We also noted this in the figure legend to fully address the reviewer's concern.

Line 233: unlikely that E. coli were grown to an OD of 600, correct typo

Corrected.

Line 261 and 285: indicate amount of TEV protease used in % w/w, also indicate temperature of digestion

We added this.

Line 68: authors should be clearer in their reference to flexible nsp8 regions as they used one of the cited papers to dock into their EM density.

We added a sentence to explain this better. Briefly, it is important to say that we know these elements are flexible because they were previously shown by others to be able to adopt different structures. We have added this. Also, we added an ED figure panel depicting nsp8 structural variability as observed in our structure by comparison of the two nsp8 copies observed (ED Figure 3c). Compare also reviewer #3, comment 3.

Line 72: indicate upon first mention which protein K58 is in and specifically what mutation is lethal.

We added this information to the text.

Line 310: indicate length of gel

We added this information.

Line 325: indicate final protein concentration prior to freezing

We were technically not able to reliably determine concentration at this particular step. However, we now provide the absorption measurement for peak protein fractions (see sentence before) and this allows for reproducing the cryo-EM sample.

Impressive RNA density

We thank the reviewer. We were also very pleased such density can be obtained provided that the complex is rather small for single particle analysis. We have added an additional ED figure panel to show the density for the first turn of the RNA (ED2f).

Model: there is no confidence in the sequence assignment or build for 2-30 of chain B. In the composite map used for refinement there is almost no density until residue 74 and this region of chain B only become visible upon low pass filtering. Similarly for chain D, there is no confident side chain density before residue 40. The same can be said for the N-terminal region of chain A.

With respect to the N-terminal regions of the chain B and D, we are confident in our modeling, because these regions were previously observed in high-resolution crystal structures, as mentioned in the Methods section, and were placed into our fine low pass filtered density that the reviewer has also examined. We made sure this is correctly described and extended the Methods accordingly. With respect to missing side chain density, this is expected for more mobile protein regions, as the reviewer surely appreciates, and we made sure we do not overinterpret potential contacts of these residues (compare our changes to Fig. 3). With respect to the N-terminal region of chain A, we do not interpret any structural details here. Again, we are sure the course of the backbone is correct.

There is also a known 9 aa register shift in the C-terminus of nsp12 (Tristan Croll describes it on Twitter, @CrollTristan) that the authors should fix in their model.

As was mentioned in the methods section, we had fixed one register shift in the nsp12 C-terminal region and had acknowledged Tristan Croll for pointing this out on Twitter. We have now also fixed the register shift at the C-terminus of nsp12 and thank the reviewer for pointing this out. The files that will be published in the PDB on May 13, 2020, are final and contain these last remaining corrections.

The supplementary movies do not add to the data or conclusions already presented.

We agree but still wish to keep these videos for general use in teaching and to allow non-experts to appreciate the three-dimensional features of the structure. We trust the reviewer agrees this is important, in particular with respect to the general interest in SARS-CoV-2.

Referee #3 (Remarks to the Author):

The short manuscript by Hillen et al describes a single particle cryoEM analysis of SARS-CoV-2 nsp12-nsp7-nsp8 polymerase complex in complex with dsRNA that mimics a replication complex.

The structure reported shows a 1:1:2 complex of nsp12-nsp7-nsp8 with the core at an average resolution of 2.9 Å. Focused refinement reveals the structure (at an average resolution of 3.5 Å) of the long helical N-terminal extensions of each of the nsp8 molecules contacting a pseudo-continuous dsRNA extending ~30 base-pairs away from the polymerase active site. The occurrence of this long RNA duplex appears to be fortuitous, resulting from the concatenation of duplexes of the input template whose sequence is partially palindromic. No NTPs were added to the sample so there has been no active RNA synthesis and the +1 site for the incoming nucleotide is vacant.

There are two recently published cryoEM structures of the SARS-CoV-2 nsp12-nsp7-nsp8 polymerase complex, the apo-structure by Gao et al (Science 2020) and the product-template complex with

incorporated remdesivir monophosphate (nominally at 2.5 Å resolution) by Yin et al (bioRxiv at the time of writing of the manuscript, now in Science since 1 May 2020).

The added value of the Hillen et al work is the first visualisation, in the context of the polymerase complex, of the helical extensions of both nsp8 molecules (these extensions are disordered in the other polymerase complex structures) which run parallel to the outgoing RNA duplex making numerous backbone contacts via mainly basic residues to both strands. Amongst these contacting residues are nsp8 K58 whose mutation to alanine has previously been shown (Subissi et al, 2014) for the highly homologous SARS-CoV-1 system to seriously affect the processivity of the polymerase. This supports the interpretation of Hillen et al that the helical extensions act as electrostatic guide through which the product template RNA duplex is extruded and that help the polymerase to grip the RNA beyond which is possible for the core nsp12 polymerase thus promoting processivity.

The paper is clearly written although some points need more discussion (see below). The methods are particularly well explained. The figures and movies are good.

In the light of the current high interest in fully understanding the mechanism of the SARS-CoV-2 polymerase and the fact that this work was done independently and in parallel with the other recently published work, this novel insight makes the paper worth publishing in Nature. However, the following points need to be taken into account.

We thank the reviewer for the fast and positive review.

(1) The title is misleading. This is not the structure of an actively replicating polymerase since no RNA synthesis has occurred. It is more correct to say it is mimicking an actively replicating polymerase

To address the reviewer's concern, we changed the abstract to clearly say that our structure shows the RdRp "…in active form, mimicking the replicating polymerase.".

On the other hand, the use of 'replicating' does not necessarily imply that the RNA was made by the enzyme. For example, many laboratories have referred to transcription complexes that contain template DNA and product RNA as 'transcribing' complexes, even if the product RNA strand was synthetic and not produced by the polymerase before structural analysis. We therefore wish to keep the title, and thereby also keep it short. Note please that we considered alternative titles but they may also be misleading. For example, "Structure of SARS-CoV-2 polymerase replication complex" first appears to be better, but could lead to the misunderstanding that we also have other nsp proteins in the complex that are additionally required, such as the capping enzyme subunits. We trust the reviewer understands our decision.

(2) To emphasise point (1), the authors need to explain in the main text the nature and origin of the dsRNA in their structure and include Ext. Data Figure 1c in the main Figure 2 or 3. As it stands, Fig 3a is misleading as it suggests the RNA is continuous which does not square with the breaks in the RNA backbone visible in the structure figures.

We now indicate the nature of the RNA template in the main Fig. 3 as suggested by the reviewer, to prevent such misunderstandings. In the figure legend, we added that "The RNA is assembled from overlapping short RNA oligos" and refer the reader to ED Figure 1c for a detailed view of the scaffold.

(3) The authors should include a figure superposing nsp8a and nsp8b showing the asymmetry.

We added a figure panel showing the superposition of nsp8a and nsp8b using their head domains (ED Figure 3c). In the main text, we refer more clearly to this asymmetry and the new figure panel.

(4) The authors show multiple basic residues from the nsp8 helices interacting with the RNA. How can they explain that mutating only K58 has such a big effect.

Unlike most other residues along the nsp8 helices, K58 projects into the RNA duplex minor groove (compare Fig 3c). Because the minor groove has very high charge density from the backbone phosphates, K58 might for this reason contribute strongly to stabilization and processivity. We added the information to the text that it binds the minor groove, but do not wish to go into any speculation. Answering the question would require extensive mutational analysis of the system that is beyond the scope of our work.

(5) Subissi et al, 2014 showed that the exonuclease-methyltransferase nsp14 protein binds to the nsp12-nsp7-nsp8 polymerase complex (for SARS-CoV-1). The authors should mention this and comment on how proof reading might occur if during processive elongation, there is such a long double stranded region.

We now mention proofreading in the main text in the last paragraph. Since this is very important in the context of understanding why remdesivir can inhibit viral replication, we included this sentence at the end of the paragraph on remdesivir. Although this must be studied in the future, we predict that for proofreading to occur, nsp14 needs to gain access to the misincorporated nucleotide at the RNA product 3'-end and catalyze its exonucleolytic removal. Steric considerations argue that this requires some rearrangements that cannot be predicted. This is why we can not comment on it now. Note that proofreading may occur in the presence of the exiting RNA duplex.

(6) Do the authors believe that product and template form a complete dsRNA over their whole length or is there a mechanism to separate the strands at some point. The authors should discuss how their observation fits into the overall mechsim of genome replication.

To our knowledge the overall mechanism of coronavirus genome replication is unclear. We assume the produced RNA duplex is converted into single strands, maybe with the action of a helicase, but we have not found literature on this. This thus seems to be an important open question. We therefore only added a little bit of text to address this point. We also briefly mention the open question of how the enzyme switches to transcription mode, when strand separation must occur to produce single-stranded viral mRNAs that can be translated.

(7) The authors use CTF refinement to improve the final cryo-EM density. Could they further improve the density by microscope optical aberration corrections implemented in RELION 3.1?

The phase residual plot generated during CTF refinement showed absence of higher-order aberrations. Therefore this was not an option to improve the density further.

(8) The authors use unsupervised 3D classification to identify more defined 'sliding poles' class. Could focus classification (perhaps without alignment) on the same region reveal an even better defined class?

Focused classification with and without alignment was indeed tested and performed using masks of varying size around the region of interest, but none of these attempts yielded better classes.

(9) Line 356 claims extraction at 1.3 Å/px, the final table 0.834 Å/px

We are sorry for this misunderstanding. We have now specified in the table that the sliding pole map (but not the core map) was refined using data binned to 1.3. This explains the difference in values that the reviewer noted.

Reviewer Reports on the First Revision:

(Referee #2)

All of my comments have been adequately addressed.

(Referee #3)

I have reviewed the revised version of the manuscript and the replies to the referees comments and am satisfied that the paper is now ready for publication.