THE EMBO JOURNAL

The CRL4DCAF1 cullin-RING ubiquitin ligase is activated following a switch in oligomerization state

Nicolas Thomä, Weaam Mohamed, Andreas Schenk, Georg Kempf, Simone Cavadini, Anja Basters, Alessandro Potenza, Wassim Abdul Rahman, Julius Rabl, and Kurt Reichermeier
DOI: 10.15252/embj.2021108008

Corresponding author(s): Nicolas Thomä (Nicolas.Thoma@fmi.ch)

Review Timeline:

Submission Date:	22nd Feb 21
Editorial Decision:	29th Mar 21
Revision Received:	16th Aug 21
Editorial Decision:	31st Aug 21
Revision Received:	7th Sep 21
Accepted:	10th Sep 21

Editor: Hartmut Vodermaier

Transaction Report:

(Note: With the exception of the correction of typographical or spelling errors that could be a source of ambiguity, letters and reports are not edited. Depending on transfer agreements, referee reports obtained elsewhere may or may not be included in this compilation. Referee reports are anonymous unless the Referee chooses to sign their reports.)

1st Editorial Decision 29th Mar 2021

Thank you for submitting your manuscript on CRL4-DCAF1 structure and regulation for our editorial consideration. It has now been assessed by three expert referees, whose comments are copied below. All reviewers find your results and conclusions interesting and potentially important, but at the same time also raise a number of concerns that would need to be addressed before publication may be warranted. As you will see, key issues relate to the cryo-EM reconstructions and their integration with the complementary biophysical/biochemical data and modeling/docking approaches. Furthermore, the referees criticize parts of the downstream functional analyses as well as the overall presentation of the experiments and data in both text and figures.

Should you be able to adequately address these key points, we would be interested in pursuing a revised manuscript further for EMBO Journal publication. Since it is our policy to consider only a single round of major revision, it will however be crucial to comprehensively respond to all the points raised at the time of resubmission. In light of the present pandemic-related disturbances and their affect on lab work, I would be open to extending the revision period beyond the default three-months time frame if needed to carefully complete such revision; with our 'scooping protection' (meaning that competing work appearing elsewhere in the meantime will not affect our considerations of your study) remaining valid also during an extended revision period.

REFEREE REPORTS

Referee #1:

In this manuscript, the authors report a low resolution structure of the CUL4-RBX1-DDB1-DCAF1 E3 ligase complex. They find that the complex is arranged as a dimer of dimers that is inactive

because the E2-binding surface on RBX1 is occluded. Neddylation activates the complex by dissociating the tetrameric complex into two dimers, and releasing RBX1 so that it can interact with E2.

Overall, this is a very interesting manuscript and it provides important new insight into this CRL4 complex and its regulation. There are many presentation issues that need to be addressed and in many places, there is a lack of clarity. Figures often do not show what I expected, making for a frustrating read.

Main comments:

1. The authors need to improve the figures. For example on line 140, Figure 2C is referenced for the DCAF1-WD40 interaction with the cullin CTD and RBX1. Figure 2C does not show this and I am unable to locate any figure that shows this clearly. Figure 3A partially shows this but the presentation needs to be improved - eg it is not obvious whether RBX1 interacts with the side of the WD40 or on the top/bottom. Figure 2C should be referenced on line 143.

2. Similarly, on lines 150 - 165 the structure of the LisH domain is described. Here, a figure would help enormously. The model is not interpretable in Figure 2C. This paragraph could be rewritten for clarity: I do not know what a 'confidence score of 0.67' means (line 157). Please rewrite 'showed convergence towards the LisH homodimer homology model' (line 160) as it is not clear what the authors are trying to say. Please explain the difference between 'symmetric docking simulations' and 'tentatively docked into the putative LisH density'. Helix 3 is shown in Figure 2C but the text states that this was not modelled due to uninterpretable density (line 165).

3. Figure S2D is referenced for the RMSD between the two ARM models but this is not found there. On lines 188-192 the authors state that they could not unambiguously dock the models into the map. Would it then be reasonable to remove the residue numbers from the docked models in Fig 2C?

4. In the paragraph starting on line 194, the authors describe how flexible fitting was used to fit the ARM in to the map. Are the helices well resolved in the map? If not, flexible fitting is likely to overinterpret the data. This is a difficult resolution to interpret and the authors should be cautious. Of course there is confidence in the overall arrangement due to previous crystal structures. The entire section on the ARM domain should be rewritten in a more cohesive way.

5. Can the authors explain how neddylation inhibits tetramer formation?

6. Paragraph lines 255-261 does not make sense. The final sentence seems to repeat what was said above.

7. Is there any indication that the dimer-tetramer transition would be regulated?

8. Please show FSC curves for the EM maps.

Minor comments

Line 89 - I do not know what the SECexplorer web platform is.

Line 131 - should be 8.4 Å resolution cryoEM map of ...

Line 146 - I think tetramerization is mediated by the DCAF1 WD40 but that is not listed. How is the cullin full-length required? It would be better to rewrite this sentence to state that tetramerization is 'mediated by interactions between'.... (instead of 'requires')

Figure 2C - please clarify the numbering of the ARM repeats. Apparently 4 ARM repeats were modelled starting with aa507 (line 172) but there are 5 in the figure starting with aa507.

Line 198 - add a reference to Fig 2C?

Please state in the main text that the cryoEM sample was crosslinked.

Figure 1C - what is the identity of the major band in the middle of the gel? Please include sizes of the markers.

Fig S2A - please show a scale bar on the micrographs

Referee #2:

Mohamed et al. contribute a cryo-EM reconstruction of the $CRL4^{DCAF1}$ E3 ubiquitin ligase at intermediate resolution. By rigid-body fitting of CUL4/RBX1, DDB1 and DCAF1-WD40 crystal structures, they generate a molecular model comprising two elementary dimers, whose dimerisation is mediated by the DCAF1 LisH domain. By antiparallel interaction of the CUL4 C-termini and RBX1 subunits of one such dimer with the DCAF1-WD40 domains of the opposing elementary dimer, a tetrameric arrangement is assembled. This molecular architecture is corroborated by SEC-MALS data and site-directed mutagenesis. In addition, docking of *in silico* models of the DCAF1 LisH and ARM-repeat domains is attempted. Based upon further molecular modelling, biochemical analysis and low-resolution cryo-EM, the authors propose a model, where the tetramer represents an auto-inhibited state of the ubiquitin ligase, which is refractory to regulation by (de-)neddylation and substrate binding. The findings provide novel and unanticipated structural and functional information on the role of DCAF1 in the context of the CRL4 complex. The results are highly relevant for the readership of EMBO Journal not only because of the essential function of DCAF1 in a variety of cellular processes, but because they also demonstrate a novel mechanism of CRL4 regulation. However, there are aspects of generation, interpretation and representation of the structural data, which deserve additional attention as outlined below.

Specific numbered comments

Major

1. Cryo-EM data collection and analysis

A major limitation of the study is the quality of the $CRL4^{DCAF1}$ cryo-EM reconstruction. Only 1280 micrographs were processed, yielding only ~160.000 particles. I'd highly recommend to collect more data (>500.000 particles) which might significantly improve downstream processing and resolution. Also, DSS cross-linking followed by SEC could be tried to improve sample quality instead of Grafix, if native sample is not suitable for plunge-freezing. In addition, the data processing scheme (Fig. S2) is

somewhat odd: after 2D classification, 3D classification was performed, but all resulting classes were pooled and then 2D-classified again. What was the reason for that procedure? What was the reason for only including classes I and IV of the second 3D classification in the final refinement? Classes II and III look similar. Lastly, if you look at classes III and IV, it seems that there might be a symmetry mismatch. Has it been tried to mask out only one dimer (either the elementary LisH or the antiparallel CUL4 dimer), and align all particles using these masks without imposing symmetry? This might at least improve the resolution of one of the dimers, and might provide valuable information on the plasticity of the binding interfaces between the dimers.
Finally, a data processing scheme for the $CRL4^{DCAF1}$/CSN complex is missing, and FSC curves for the final refinement of both datasets.

2. Docking of in silico models of DCAF1 N-terminal domains
It is very hard to gauge the quality of fit of the docked homology models from Fig. 2. At the very least, I'd expect to see enlarged supplementary figures of the corresponding density segments together with the fitted models to see if there is actually tubular density corresponding to the fitted helices. The present figures do not look very convincing. Also, why is only a small part of the ARM domain homology model fitted, even if that model seems to form one stable folded unit (Fig. S2C)? The authors should provide additional experimental evidence, i.e. better maps (see point 1), or supporting analyses such as cross-linking mass spectrometry. Otherwise, I'd refrain from modelling these regions, which would, in my opinion, not influence the general conclusions drawn from the structural analysis in any way.

3. Auto-ubiquitylation assay
The auto-ubiquitylation activity is crucial experimental evidence to support the auto-inhibition hypothesis, i.e. that the tetrameric assembly sterically blocks the access of ubiquitin-charged E2 to RBX1 (Fig. 3C). Accordingly, a control demonstrating equal loading of the compared blots should be included. Furthermore, it would be interesting to know which $CRL4^{DCAF1}$ component is actually auto-ubiquitylated. In addition, kinetic analysis of Vpr-induced UNG2 ubiquitylation with WT and R1247A $CRL4^{DCAF1}$ could be performed as an example of substrate ubiquitylation.

Minor

1. Line 117 "Observed MW 838 kDa": in Fig. 1B it says 780 kDa
2. Lines 138-140 "The $CRL4^{DCAF1}$ map indicates ... interaction between the DCAF1 (WD40) ..., with the cullin CTD and RBX1 ... (Fig. 2C)": in Fig. 2C, DCAF1 (WD40) is not indicated at all. It rather looks like there are interactions between DCAF1 ARM domain and RBX1 or DDB1 BPA and CUL4-CTD. Could this be clarified?
3. Line 146: "tetramerization requires cullin full length": have any cullin truncations been tested?
4. Line 228: "inactive": less active?
5. Line 334: "receptor auto-ubiquitination": see major point 3, is it actually the DCAF1 receptor which is auto-ubiquitinated?
6. Line 342: "CUl1" should read CUL1
7. Line 360: which volumes and relative ratios of baculoviruses have been used for co-infection per which volume of cells?
8. Lines 367, 373: Which concentration of Tris-HCl?
9. Lines 395-398: Please state protein concentrations, bead and buffer volumes used in pull-downs
10. Lines 402-408: Please state protein volumes and concentrations used for SEC
11. Lines 427, 447: Which kind of grid has been used for the respective data collection? R1.2/1.3 or

lacey with thin carbon support?
12. Lines 462-463: Why have the crystallographic models been re-refined?
13. Some references seem incomplete (e.g. Hrecka 2007, Marks 2016, Rohou 2015, Stark 2010, Wang 2016, Yang 2020, Yu 2013, Zhang 2001)
14. Fig. 5B is not referenced in the text
15. Fig. 5C and S5C are mis-labelled: UNG2 is actually Vpr and vice versa
16. Fig. S1A: please explain "sensitivity score" in the legend
17. Figs. S1B, D, S2A, B, C, are not referenced in the text
18. Fig. S4A left panel: control trace (N8-CRL4DCAF1 FL) is missing

Referee #3:

Mohamed et al. reported biochemical and structural characterization of Cul4-Rbx1-DDB1-DCAF1 complex and revealed an autoinhibited tetrameric configuration that is regulated by Nedd8 modification and substrate binding. SEC-MALS experiment and a low-resolution Cryo-EM model confirmed a tetrameric configuration of Cul4-Rbx1-DDB1-DCAF1 complex. Prior study showed that LisH motif in DCAF1 could form a dimer. With this knowledge, the authors docked the modelled DCAF1 armadillo and LisH structures into the density. The model shows that LisH domain forms a dimer and the dimer further assembles into a tetramer via intermolecular Rbx1 and DCAF1 WD40 domain interaction. A VRSA-loop in the WD40 domain contacts Rbx1 RING domain. Alanine substitution of the conserved R1247 in this loop disassembled the tetramer into dimer confirming the importance of this interaction. The tetrameric configuration is not compatible with Ub-loaded E2 or Nedd8-loaded E2 binding. The authors showed that WT-complex had reduced autoubiquitination activity and Ubc12-mediated Nedd8 modification of Cul4 compared to R1247A-complex suggesting that tetramer represents the inactive conformation. Prior studies showed that Nedd8 modification of Cullin induces conformational changes in the C-terminus of Cullin and Rbx1; the conformational change would abolish Rbx1 and WD40 interaction. Consistent with this notion the authors showed that Nedd8-modified complex exists as a dimer. Moreover, the tetrameric model is incompatible with CSN interaction and the authors showed that CSN did not pull down tetrameric form of the CRL4-DCAF1 complex. Lastly, the authors showed that substrate, Vpr-UNG2, binding to the WD40 domain of DCAF1 converts the complex into a dimer, whereas non-substrate binding such as MERLIN had no effect on the tetrameric assembly.

This study reveals a novel autoinhibited tetrameric configuration of CRL4-DCAF1 complex where Rbx1 is partially occluded from binding to E2-Ub or E2-Nedd8. The autoinhibition is released when CRL4 is Nedd8 modified or when DCAF1 engages a substrate. These findings will be of interest to EMBO Journal readers.

I have few comments that need to be addressed.

1. The authors mentioned that CRL4-DCAF1 complex exists in a dimer/tetramer equilibrium, but SEC-MALS shows a single mono-disperse species. What is the protein concentration of CRL4-DCAF1 complex used in the SEC-MALS experiment? Is the tetramer formation concentration dependent?

In the Cryo-EM analysis, is the complex predominantly tetramer or is there a small fraction that is dimer?

2. The model showed that the tetrameric configuration is incompatible with E2 binding, but CRL4-

DCAF1 is active in ubiquitination and neddylation suggesting that it can still bind E2. The authors suggest the CRL4-DCAF1 is in a dimer/tetramer equilibrium and it is the dimer population that is active. It is unclear from the data whether the complex exists in a dimer/tetramer equilibrium. Could addition of Ub-loaded E2 or Nedd8-loaed E2 compete with WD40-RING interaction in the tetramer and cause the disassembly of tetramer? In a way, the tetrameric configuration reduces the binding affinity for Ub/Nedd8-loaded E2?

3. In Figure 4B the loading of middle two panels are not the same. CRL4-DCAF1 (R1247A) bands seem less intense compared to WT panel. It is evident that there is a faint un-neddylated CRL4 R1247A band, which will show up if loaded similarly as WT panel. It would be useful to compare the initial rate of Nedd8-CRL4 formation using shorter reaction time point since at 5-min time point neddylation is nearly complete.

4. In Figure 5A and 5D, CRL4-DCAF1-VPR, CRL4-DCAF1-VPR-UNG2 and CRL4-DCAF1-MERLIN elution profiles have a large peak near the void volume. Do these complexes form larger oligomer? Please explain.

5. In Figure S4 why is initial rate Vo expressed as uM/s? This is different from kcat (s-1).
In Figure S4B, the y-axis numbers are mislabelled? kcat is ~0.6 s-1 but the curve is showing 0.006.

Point-by-point letter highlighting our response to all comments/suggestions:

Referee #1:

In this manuscript, the authors report a low resolution structure of the CUL4-RBX1-DDB1-DCAF1 E3 ligase complex. They find that the complex is arranged as a dimer of dimers that is inactive because the E2-binding surface on RBX1 is occluded. Neddylation activates the complex by dissociating the tetrameric complex into two dimers, and releasing RBX1 so that it can interact with E2.

Overall, this is a very interesting manuscript and it provides important new insight into this CRL4 complex and its regulation. There are many presentation issues that need to be addressed and in many places, there is a lack of clarity. Figures often do not show what I expected, making for a frustrating read.

Main comments

1. **The authors need to improve the figures. For example on line 140, Figure 2C is referenced for the DCAF1-WD40 interaction with the cullin CTD and RBX1. Figure 2C does not show this and I am unable to locate any figure that shows this clearly. Figure 3A partially shows this but the presentation needs to be improved - eg it is not obvious whether RBX1 interacts with the side of the WD40 or on the top/bottom. Figure 2C should be referenced on line 143.**

   We apologize that we mistakenly referenced **Figure 2C** to show DCAF1 WD40 interaction with the cullin CTD and RBX1. This is now referenced correctly to Figure 3A (line 151), and we added an additional panel to detail the DCAF1 WD40 interaction with the Cullin CTD and RBX1 (**Figure 3A**). Structure figures have been improved throughout, for example, Figure 2C has been moved to an Extended View figure (**Figure EV 3A**), as we dedicated this figure now to better describe the LisH domain and ARM repeat structures (as explained in details below).

2. **Similarly, on lines 150 - 165 the structure of the LisH domain is described. Here, a figure would help enormously. The model is not interpretable in Figure 2C. This paragraph could be rewritten for clarity: I do not know what a 'confidence score of 0.67' means (line 157). Please rewrite 'showed convergence towards the LisH homodimer homology model' (line 160) as it is not clear what the authors are trying to say. Please explain the difference between 'symmetric docking simulations' and 'tentatively docked into the putative LisH density'.**

   We agree that the description of the modelling and the fitting of the LisH domain to the map needed improvement. An additional supplementary figure is now included to detail the LisH domain architecture as well as the ARM repeat structure (**Figure EV 3A and D**) and the quality of the fit of the model to the density. We also re-wrote the text and methods sections to make the procedure more transparent, and now include a reference to better explain the confidence score (line 171). The paragraph now reads: “The putative LisH density proximal to the WD40 domain showed helical features and strongly suggested the presence of a two-fold symmetry axis in line with previously observed

LisH domain homodimers (PDB 1UUJ, PDB 1VYH) (Kim *et al*, 2004; Tarricone *et al*, 2004). The LisH domain fold is characterized by a two-helix bundle (helices 1 and 2) with a third helix (helix 3) crossing the helical bundle (e.g. PDB 1UUJ, PDB 6IWV). Upon homodimerization, the two-helix bundles form a four-helix bundle in which helices 1 and 2 align in an anti-parallel and slightly diagonal manner, respectively. In several structures, helices 3 of the homodimeric assembly additionally align in a diagonal manner perpendicular to the four-helix bundle (PDB 1UUJ, PDB 6IWV). The DCAF1 LisH domain structure was predicted by comparative modelling with high confidence (confidence score of 0.67, Song et al, 2013). To obtain a model for the homodimeric complex, the monomeric consensus model from comparative modelling (aa 846-883) was superposed on the dimeric LisH domains from PDB 6IWV. The homodimer interface was independently validated by docking two separated LisH monomer models against each other, imposing two-fold symmetry constraints (see Methods). In these simulations, the LisH dimer interface previously observed in homologous dimeric LisH crystal structures gave the best docking scores (**Figure EV 3B**). The dimeric homology model (obtained from superposition on a template structure) was then docked into the putative LisH density (**Figure EV 3A and D**). The density supports an anti-parallel alignment of helices 2, while the putative density for helices 1 is fragmented. We observed only uninterpretable density at the expected location for helices 3 and refrained from modeling this helix given the limited local resolution.

A significant portion of uninterpreted density located between DDB1-BPA and DDB1-BPB of different protomers and close to the putative LisH density showed features indicative of several α-helical bundles. These features would be in agreement with an armadillo fold predicted for the segment N-terminal to the DCAF1 LisH domain (ARM, 1-817). To obtain a model for this part, we employed the deep learning-based structure prediction pipeline *Alphafold* that has been shown to yield highly accurate predictions even in case of targets where no structural templates of close homologs are available (**Figure EV 3C**; (Jumper *et al*, 2021)). The part directly N-terminal to the LisH domain (aa 507-817), which was approximately matching the volume of the remaining density, was extracted from the top-ranked model of *Alphafold*. In the predicted model, the extracted portion contains four complete armadillo (ARM) repeats each composed of a short helix that is perpendicular to two somewhat diagonal and anti-parallel aligned longer helices and an incomplete repeat (third helix only) at the N-terminus. This particular helix was predicted to be significantly longer. Due to the strong repetitiveness of the α-helical repeats and the poor map quality in this region, it was not possible to unambiguously dock the model and we cannot exclude that the observed density corresponds to a different part of the ARM domain. In Figure EV 3A the three best fitting poses with real-space correlations between 0.52 and 0.59 (calculated with a simulated model-map at 8.4 Å) are shown.

In our fitted models, a gap of around 33-46 Å with uninterpretable density remained between the C-terminal part of the DCAF1 ARM domain and the DCAF1 LisH domain (**Figure EV 3A, D**). These distances would be all consistent with the length of a linker region of 29 aa, which is predicted to be (partially) α-helical. Some ARM units are interspersed by loop regions and, in particular, the four C-terminal repeats are connected to the next N-terminal repeat by a region that has been previously suggested to contain a

small chromo-like domain (Schabla *et al*, 2018). Owing to the limited local resolution, the presence of a chromo-like fold could not be verified. No density was observed for the remaining N-terminal part of the ARM domain that would be consistent with an interspersed casein-kinase like domain (Kim *et al*, 2013a).

Overall, the density suggests the presence of a symmetric α-helical interaction motif and several armadillo-like repeats in close proximity that likely correspond to the LisH domain and the C-terminal part of the ARM domain, respectively. Predicted models for these regions of the map can largely explain the density features at the given resolution. Higher resolution experimental data is required to unambiguously dock a model and carry out more detailed residue assignments."

Helix 3 is shown in Figure 2C but the text states that this was not modelled due to uninterpretable density (line 165).

We also agree that adding residues at the end of helix 3 is confusing given the limited resolution, and we have now removed these.

3. **Figure EV 2D is referenced for the RMSD between the two ARM models but this is not found there. On lines 188-192 the authors state that they could not unambiguously dock the models into the map. Would it then be reasonable to remove the residue numbers from the docked models in Fig 2C?**

   The ARM model (now predicted only with Alphafold v2.0) is shown in a new supplementary figure (**Figure EV 3C**). Following the suggestion from the reviewer, we removed the numbers of the helices and moved the panel (**Figure 2C**) to a supplementary figure **(Figure EV 3A**). We also now provide a more detailed explanations for the figure and methodology used in the corresponding methods section and figure legend.

4. **In the paragraph starting on line 194, the authors describe how flexible fitting was used to fit the ARM in to the map. Are the helices well resolved in the map? If not, flexible fitting is likely to overinterpret the data. This is a difficult resolution to interpret and the authors should be cautious. Of course there is confidence in the overall arrangement due to previous crystal structures. The entire section on the ARM domain should be rewritten in a more cohesive way.**

   Following the suggestion of the reviewer, we fully rewrote the section in question and now also added new figures (**Figure EV 3A and D**) to show the quality of the fit to the density. We agree that there is no clear helical density, however, there are features visible in the 8.4 Å map reminiscent of helical bundles. We now show different potential docking poses for the ARM domain that have similar map-model correlations to highlight the ambiguity.

   The map interpretation is based on the density, sequence and modelling data, and although we do not plan on submitting this part of the model as we cannot assign individual residues, we think that the figures and the overall map interpretation add to the paper and thus prefer to keep them in the manuscript.

5. **Can the authors explain how neddylation inhibits tetramer formation?**

   Following the reviewer's question, we added a possible model discussing this point in the discussion section (line 346). The text now reads: "Several studies have suggested that the cullin CTD is flexible upon neddylation (Duda *et al*, 2008; Baek *et al*, 2020; Angers *et al*, 2006; Fischer *et al*, 2011; Banchenko *et al, 2021*). These conformational changes triggered upon neddylation in turn may also alter the interface between DCAF1 WD40 and RBX1. Unlike recent findings for the CUL1 systems (Baek *et al*, 2020), it is currently unclear how CUL4 ligases interact with ubiquitin loaded E2 enzymes. Given the extensive conformational rearrangements observed for CUL1 CTD, however, it is likely that a neddylated $CRL4^{DCAF1}$ in its E2-ubiquitin bound form would further impair tetramer formation in cells. Moreover, although our findings are consistent with $CUL4^{DCAF1}$ being in a dimer-tetramer equilibrium, we cannot exclude that Ub/Nedd8-loaded E2 complexes, or in fact also CSN, possess additional molecular mechanisms to open the $CUL4^{DCAF1}$ tetrameric complex and thereby convert it to a dimer."

6. **Paragraph lines 255-261 does not make sense. The final sentence seems to repeat what was said above.**

   We apologize for this oversight and have adjusted the text accordingly (line 273).

7. **Is there any indication that the dimer-tetramer transition would be regulated?**

   This is a very good question. We have extended our discussion to include our hypothesis on this critical point. This has now been added (line 376) as follows:" We speculate that the $CRL4^{DCAF1}$ dimer-tetramer transition is primarily regulated through neddylation. However, a possible regulatory mechanism involving post-translational modifications including phosphorylation can also not be excluded at this point. There are a number of residues in the CUL4 C-terminus expected to be phosphorylated (S642, Y744, and N751), based on the Phosphositeplus resource (Hornbeck PV *et al.*, 2015). When phosphorylated, these residues have the potential to impact tetramerization in light of our structures."

8. **Please show FSC curves for the EM maps.**

   We have now included the FSC curves as suggested (**Figures EV 2C and 4E**). When generating the FSC curve for the CSN-$CRL4^{DCAF1}$ map, we observed that a C2-symmetrized map resulted in a higher quality map compared to one where no symmetry is applied (see also reviewer 2). We have thus decided to use the C2 map instead of the previous map where symmetry is applied in the revised manuscript. We thank the reviewers for pointing this out.

Minor comments

1. **Line 89 - I do not know what the SECexplorer web platform is.**

The SECexplorer data provides an apparent molecular weight for protein complexes in mammalian cells. An explanation of this web resource has now been added to the text (line 88). The text now reads: “The SECexplorer workflow fractionates native protein complexes by size Exclusion Chromatography (SEC) followed by mass spectrometry to identify proteins in each fraction, along with an apparent molecular weight for the eluting complexes (Heusel *et al*, 2019). The size exclusion profile in SECexplorer finds DCAF1 eluting with large native protein complexes of more than 1 MDa.”

2. **Line 131 - should be 8.4 Å resolution cryoEM map of ...**

   Many thanks for pointing this out, we have now adjusted the text accordingly (line 141).

3. **Line 146 - I think tetramerization is mediated by the DCAF1 WD40 but that is not listed. How is the cullin full-length required? It would be better to rewrite this sentence to state that tetramerization is 'mediated by interactions between'.... (instead of 'requires')**

   We edited the text according to the reviewer suggestion (line 155). The cullin 4 full-length protein encompasses the N-terminal domain (NTD), which binds DDB1, and the C-terminal domain (CTD) with its RBX1 subunit. DDB1 and RBX1 are required for DCAF1 binding and tetramerization. The text now reads as: “The structural data is consistent with the results obtained by SEC-MALS, illustrating that $CRL4^{DCAF1}$ is tetrameric, and that tetramerization is mediated by the interactions of the cullin CTD-RBX1 and DCAF1 WD40 domain at one dimerization interface, and the DCAF1 N-terminal LisH motifs at the other.”
   We thank the reviewer for helping us to bring across this point more clearly.

4. **Figure 2C - please clarify the numbering of the ARM repeats. Apparently 4 ARM repeats were modelled starting with aa507 (line 172) but there are 5 in the figure starting with aa507.**

   We apologize for this oversight. There are indeed four complete ARM repeats. The fifth ARM repeat is incomplete, and therefore the figure labeling contains 5 ARM repeats.

5. **Line 198 - add a reference to Fig 2C?**

   We have moved the panel in Figure 2C to a new supplementary figure (**Figure EV 3A**). The referencing has now been fixed in the text (line 201).

6. **Please state in the main text that the cryoEM sample was crosslinked.**

   This information has now been included in the main text (line 138) as follows: “In an effort to increase the stability of $CRL4^{DCAF1}$ protein sample for freezing cryo-EM grids, we had to cross-link the sample using a gradient fixation protocol (GraFix) (Stark, 2010).”

7. **Figure 1C - what is the identity of the major band in the middle of the gel? Please include sizes of the markers.**

We thank the reviewer for pointing this out, and apologize for the oversight. We adjusted the figure labeling (**Figure 1B)** and figure legend accordingly. The major band is UBC12. This gel describes the neddylation reaction of $CRL4^{DCAF1}$ at time 0 and after 2h. This information has now been included in the main text (line 117) and in the figure legend (line 855).

8. **Fig S2A - please show a scale bar on the micrographs**

A scale bar has now been added (**Figures EV 2A and 4C**), as suggested.

Referee #2:

Mohamed et al. contribute a cryo-EM reconstruction of the CRL4^{DCAF1} E3 ubiquitin ligase at intermediate resolution. By rigid-body fitting of CUL4/RBX1, DDB1 and DCAF1-WD40 crystal structures, they generate a molecular model comprising two elementary dimers, whose dimerisation is mediated by the DCAF1 LisH domain. By antiparallel interaction of the CUL4 C-termini and RBX1 subunits of one such dimer with the DCAF1-WD40 domains of the opposing elementary dimer, a tetrameric arrangement is assembled. This molecular architecture is corroborated by SEC-MALS data and site-directed mutagenesis. In addition, docking of *in silico* models of the DCAF1 LisH and ARM-repeat domains is attempted. Based upon further molecular modelling, biochemical analysis and low-resolution cryo-EM, the authors propose a model, where the tetramer represents an auto-inhibited state of the ubiquitin ligase, which is refractory to regulation by (de-)neddylation and substrate binding. The findings provide novel and unanticipated structural and functional information on the role of DCAF1 in the context of the CRL4 complex. The results are highly relevant for the readership of EMBO Journal not only because of the essential function of DCAF1 in a variety of cellular processes, but because they also demonstrate a novel mechanism of CRL4 regulation. However, there are aspects of generation, interpretation and representation of the structural data, which deserve additional attention as outlined below.

Specific numbered comments

Major

1. **Cryo-EM data collection and analysis**

   A major limitation of the study is the quality of the CRL4^{DCAF1} cryo-EM reconstruction. Only 1280 micrographs were processed, yielding only ~160.000 particles. I'd highly recommend to collect more data (>500.000 particles) which might significantly improve downstream processing and resolution.

   A large number of datasets were acquired in the course of this project and only a few of those were of sufficient quality to be used for the final reconstruction. We found the CRL4^{DCAF1} sample to be very difficult to freeze and collect. The protein complex is very fragile and sensitive to the air/water interface even in the presence of detergents and continuous carbon support. To obtain the data included here, we merged 14 days of highly selective, manual, data collection, where 6148 micrographs were collected before choosing the best 1280. We also had several data collection sessions on a K2 or Falcon 3 with samples prepared in different buffers, different purification strategies and different grids, where the complex was mostly broken, and with the sample too heterogenous for further processing. We also attempted to improve sample quality by addition of a binding partner such as MERLIN, where we manually acquired 13954 micrographs within 16 days of cryo-EM data collection. Unfortunately, none of these efforts yielded a better resolution than what is presented in this manuscript.

Also, DSS cross-linking followed by SEC could be tried to improve sample quality instead of Grafix, if native sample is not suitable for plunge-freezing.

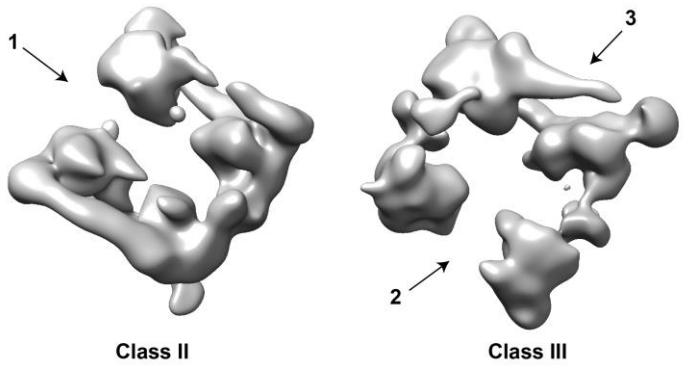
In our experience, we did not have much success with DSS crosslinkers thus far, but did not try it in this particular case. However, at the initial screening steps at the beginning of this project, we did use glutaraldyde cross-linking in-tube followed by SEC, but noticed that the sample behaved significantly better with Grafix.

In addition, the data processing scheme (Fig. S2) is somewhat odd: after 2D classification, 3D classification was performed, but all resulting classes were pooled and then 2D-classified again. What was the reason for that procedure?

We have actually excluded Class III from the subsequent analysis, and apologize that this was not made clear in our cryo-EM work-flow. We have now rectified this in the current version (**Figure EV 2B**).

What was the reason for only including classes I and IV of the second 3D classification in the final refinement? Classes II and III look similar.

Based on visual inspection of all the classes, we decided to pool Class I and IV as they looked sufficiently similar and most intact compared to the other classes. We did not include Classes II or III as they lack many structural features. Here, we include another rotational view of Classes II and III to show that in Class II the LisH dimerization motif is missing in one dimer (arrow 1), and that in Class III, two cullin arms are missing (arrow 2) and the third cullin arm is poorly-defined (arrow 3).



Lastly, if you look at classes III and IV, it seems that there might be a symmetry mismatch

We did not apply any symmetry during 3D classification, we apologize if it was not made clear. The figure in question (**Figure EV 2B**) has now been adjusted to make this clear.

Has it been tried to mask out only one dimer (either the elementary LisH or the antiparallel CUL4 dimer), and align all particles using these masks without

imposing symmetry? This might at least improve the resolution of one of the dimers, and might provide valuable information on the plasticity of the binding interfaces between the dimers.

We tried to mask out one dimer and to process it separately in the CRL4^{DCAF1}-MERLIN datasets. This workflow did not improve the resolution or the quality of the data. One key problem encountered was that Relion did not deal well with the centering, as the single dimer is off-center in the map after masking**.**

Finally, a data processing scheme for the CRL4^{DCAF1}/CSN complex is missing, and FSC curves for the final refinement of both datasets.

We apologize for this oversight and have now included the FSC curves (**Figure EV 2C, and 4E**) and the processing scheme of CRL4^{DCAF1}-CSN complex (**Figure EV 4D**). When generating the FSC curve for the CSN-CRL4^{DCAF1} map, we observed that a C2-symmetrized map resulted in a higher quality map compared to one where no symmetry is applied. We have thus decided to use the C2 map instead of the previous map where symmetry is applied in the revised manuscript. We thank the reviewers for pointing this out.

2. **Docking of in silico models of DCAF1 N-terminal domains. It is very hard to gauge the quality of fit of the docked homology models from Fig. 2. At the very least, I'd expect to see enlarged supplementary figures of the corresponding density segments together with the fitted models to see if there is actually tubular density corresponding to the fitted helices. The present figures do not look very convincing. Also, why is only a small part of the ARM domain homology model fitted, even if that model seems to form one stable folded unit (Fig. S2C)? The authors should provide additional experimental evidence, i.e. better maps (see point 1), or supporting analyses such as cross-linking mass spectrometry. Otherwise, I'd refrain from modelling these regions, which would, in my opinion, not influence the general conclusions drawn from the structural analysis in any way.**

Following the suggestions of reviewers #1 and #2, we rewrote the section in question and now also added new figure panels (**Figure EV 3A and D**) to show the quality of the fit to the density. We agree that there is no clear helical density, however, there are features visible in the 8.4 Å map reminiscent of helical modules. We now show different docking poses for the ARM domain that have similar map-model correlations to highlight the ambiguity. The section now reads: “The putative LisH density proximal to the WD40 domain showed helical features and strongly suggested the presence of a two-fold symmetry axis in line with previously observed LisH domain homodimers (PDB 1UUJ, PDB 1VYH) (Kim *et al*, 2004; Tarricone *et al*, 2004). The LisH domain fold is characterized by a two-helix bundle (helices 1 and 2) with a third helix (helix 3) crossing the helical bundle (e.g. PDB 1UUJ, PDB 6IWV). Upon homodimerization, the two-helix bundles form a four-helix bundle in which helices 1 and 2 align in an anti-parallel and slightly diagonal manner, respectively. In several structures, helices 3 of the homodimeric assembly additionally align in a diagonal manner perpendicular to the four-helix bundle (PDB 1UUJ, PDB 6IWV). The DCAF1 LisH domain structure was predicted

by comparative modelling with high confidence (confidence score of 0.67, Song et al, 2013). To obtain a model for the homodimeric complex, the monomeric consensus model from comparative modelling (aa 846-883) was superposed on the dimeric LisH domains from PDB 6IWV. The homodimer interface was independently validated by docking two separated LisH monomer models against each other, imposing two-fold symmetry constraints (see Methods). In these simulations, the LisH dimer interface previously observed in homologous dimeric LisH crystal structures gave the best docking scores (**Figure EV 3B**). The dimeric homology model (obtained from superposition on a template structure) was then docked into the putative LisH density (**Figure EV 3A and D**). The density supports an anti-parallel alignment of helices 2, while the putative density for helices 1 is fragmented. We observed only uninterpretable density at the expected location for helices 3 and refrained from modeling this helix given the limited local resolution.

A significant portion of uninterpreted density located between DDB1-BPA and DDB1-BPB of different protomers and close to the putative LisH density showed features indicative of several α-helical bundles. These features would be in agreement with an armadillo fold predicted for the segment N-terminal to the DCAF1 LisH domain (ARM, 1-817). To obtain a model for this part, we employed the deep learning-based structure prediction pipeline *Alphafold* that has been shown to yield highly accurate predictions even in case of targets where no structural templates of close homologs are available (**Figure EV 3C**; (Jumper *et al*, 2021)). The part directly N-terminal to the LisH domain (aa 507-817), which was approximately matching the volume of the remaining density, was extracted from the top-ranked model of *Alphafold*. In the predicted model, the extracted portion contains four complete armadillo (ARM) repeats each composed of a short helix that is perpendicular to two somewhat diagonal and anti-parallel aligned longer helices and an incomplete repeat (third helix only) at the N-terminus. This particular helix was predicted to be significantly longer. Due to the strong repetitiveness of the α-helical repeats and the poor map quality in this region, it was not possible to unambiguously dock the model and we cannot exclude that the observed density corresponds to a different part of the ARM domain. In Figure EV 3A the three best fitting poses with real-space correlations between 0.52 and 0.59 (calculated with a simulated model-map at 8.4 Å) are shown.

In our fitted models, a gap of around 33-46 Å with uninterpretable density remained between the C-terminal part of the DCAF1 ARM domain and the DCAF1 LisH domain (**Figure EV 3A, D**). These distances would be all consistent with the length of a linker region of 29 aa, which is predicted to be (partially) α-helical. Some ARM units are interspersed by loop regions and, in particular, the four C-terminal repeats are connected to the next N-terminal repeat by a region that has been previously suggested to contain a small chromo-like domain (Schabla *et al*, 2018). Owing to the limited local resolution, the presence of a chromo-like fold could not be verified. No density was observed for the remaining N-terminal part of the ARM domain that would be consistent with an interspersed casein-kinase like domain (Kim *et al*, 2013a).

Overall, the density suggests the presence of a symmetric α-helical interaction motif and several armadillo-like repeats in close proximity that likely correspond to the LisH domain

and the C-terminal part of the ARM domain, respectively. Predicted models for these regions of the map can largely explain the density features at the given resolution. Higher resolution experimental data is required to unambiguously dock a model and carry out more detailed residue assignments."

The map interpretation is based on the density, sequence and modelling data, and although we do not plan on submitting this part of the model as we cannot assign individual residues, we think that the figures and the overall map interpretation add to the paper and thus prefer to keep them in the manuscript.

Regarding the ARM repeats, we observed that there is likely some sort of a hinge after the modelled unit, and the N-terminal part might thus also be more flexible. We have seen such hinges in helical sections in a number of structures (Cavadini *et al*. 2016). Despite being a structured helical solenoid, part of the repeat is sufficiently mobile to become disordered.

3. **Auto-ubiquitylation assay:**

The auto-ubiquitylation activity is crucial experimental evidence to support the auto-inhibition hypothesis, i.e. that the tetrameric assembly sterically blocks the access of ubiquitin-charged E2 to RBX1 (Fig. 3C). Accordingly, a control demonstrating equal loading of the compared blots should be included

This point is well-taken. We have now included a blot of DDB1 (**Figure 3C**) as a control for equal loading. This control further supports our conclusions.

Furthermore, it would be interesting to know which CRL4^{DCAF1} component is actually auto-ubiquitylated.

Based on the molecular weight of the ubiquitination smear, which starts after ~150 kDa up to 250 kDa, DCAF1 is likely the auto-ubiquitinated component of the ligase.

In addition, kinetic analysis of Vpr-induced UNG2 ubiquitylation with WT and R1247A CRL4^{DCAF1} could be performed as an example of substrate ubiquitylation.

This is a great suggestion! We have carried out the experiment, and observe a more active ubiquitnation activity for the mutant as compared the wild-type DCAF1 (**Figure 3D)**, as expected (**Figure 3C**). These findings have now been added to the manuscript (line 241) as follows: "Reduced catalytic activity of the tetrameric CRL4^{DCAF1} compared to the dimeric mutant was also observed towards the viral substrate VPR-UNG2 (**Figure 3D**). On the other hand, preventing tetramer formation by incorporating the mutant DCAF1 (R1247A) into the complex overcomes auto-inhibition and gives rise to a more active E3 ligase (**Figure 3C, D**). Taken together, these data suggest that the tetrameric CRL4^{DCAF1} ligase represents a conformation with a reduced catalytic activity"

<u>Minor comments:</u>

1. **Line 117 "Observed MW 838 kDa": in Fig. 1B it says 780 kDa**

We adjusted this in line 123 accordingly.

2. **Lines 138-140 "The $CRL4^{DCAF1}$ map indicates ... interaction between the DCAF1 (WD40) ..., with the cullin CTD and RBX1 ... (Fig. 2C)": in Fig. 2C, DCAF1 (WD40) is not indicated at all. It rather looks like there are interactions between DCAF1 ARM domain and RBX1 or DDB1 BPA and CUL4-CTD. Could this be clarified?**

   We apologize that we mistakenly referenced **Figure 2C** to show DCAF1 WD40 interaction with the cullin CTD and RBX1. This is now referenced correctly to Figure 3A (line 151), and we added an additional panel to detail the DCAF1 WD40 interaction with the Cullin CTD and RBX1 (**Figure 3A**). Structure figures have been improved throughout, for example, Figure 2C has been moved to an Extended View figure (**Figure EV 3A**), as we dedicated this figure now to better describe the LisH domain and ARM repeat structures.

3. **Line 146: "tetramerization requires cullin full length": have any cullin truncations been tested?**

   We edited the text according to the reviewer suggestion (line 155). The cullin 4 full-length protein encompasses the N-terminal domain (NTD), which binds DDB1, and the C-terminal domain (CTD) with its RBX1 subunit. DDB1 and RBX1 are required for DCAF1 binding and tetramerization. The text now reads as: “The structural data is consistent with the results obtained by SEC-MALS, illustrating that $CRL4^{DCAF1}$ is tetrameric, and that tetramerization is mediated by the interactions of the cullin CTD-RBX1 and DCAF1 WD40 domain at one dimerization interface, and the DCAF1 N-terminal LisH motifs at the other.”
   We thank the reviewer for helping us to bring across this point more clearly.

4. **Line 228: "inactive": less active?**

   We adjusted the text accordingly (line 245), it reads now: “Taken together, these data suggest that the tetrameric $CRL4^{DCAF1}$ ligase represents a conformation with a reduced catalytic activity.”

5. **Line 334: "receptor auto-ubiquitination": see major point 3, is it actually the DCAF1 receptor which is auto-ubiquitinated?**

   We have adjusted the text according to the reviewer’s suggestion. It reads now as follows: “The wild-type tetrameric $CRL4^{DCAF1}$ complex shows a prominent reduction in DCAF1 auto-ubiquitination as compared to the dimeric mutant $CRL4^{DCAF1}$ (R1247A) (**Figure 3C**)”.

   Based on the molecular weight of the ubiquitination smear, which starts after ~150 kDa up to 250 kDa, DCAF1 is likely the auto-ubiquitinated component of the ligase.

6. **Line 342: "CUl1" should read CUL1.**

Many thanks for spotting this typo, it is now adjusted in line 349.

7. **Line 360: which volumes and relative ratios of baculoviruses have been used for co-infection per which volume of cells?**

   We thank the reviewer for pointing this out. This information has now been added to the materials and methods section (line 402).

8. **Lines 367, 373: Which concentration of Tris-HCl?**

   We have used 50 mM Tris-HCl in the buffers indicates. This information has now been added in lines 408 and 415.

9. **Lines 395-398: Please state protein concentrations, bead and buffer volumes used in pull-downs**

   We thank the reviewer for making us aware of this omission, the molar ratios between the protein complexes, beads and buffer volumes are now included in lines 437-442.

10. **Lines 402-408: Please state protein volumes and concentrations used for SEC downs**

    We thank the reviewer for making us aware of this omission, this has now been adjusted in line 447.

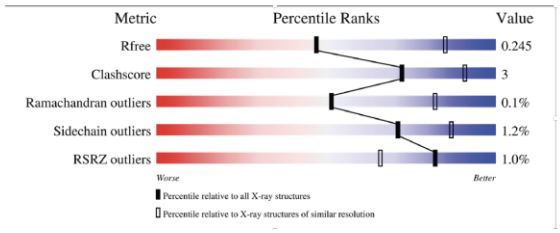
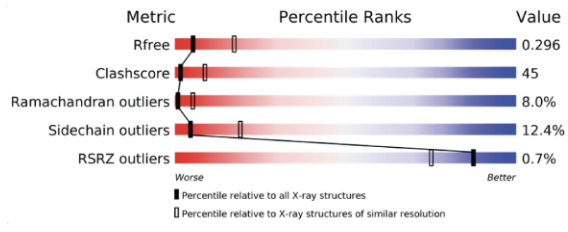
11. **Lines 427, 447: Which kind of grid has been used for the respective data collection? R1.2/1.3 or lacey with thin carbon support?**

    This information is now included in line 468. The text now reads: “For the $CRL4^{DCAF1}$ samples, R1.2/1.3, Cu 400 mesh grids were used for data collection (Quantifoil Micro Tools GmbH, Grosslöbichau, Germany). For the $CRL4^{DCAF1}$-CSN sample, Lacey carbon grids (Ted Pella, Inc) were used. For both samples, $CRL4^{DCAF1}$ and $CRL4^{DCAF1}$-CSN, grids were manually floated with continuous carbon films.”

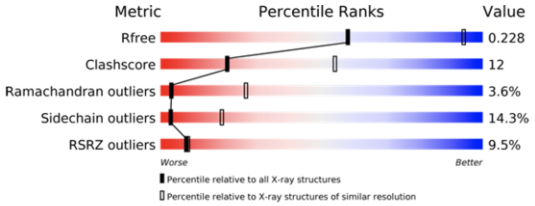
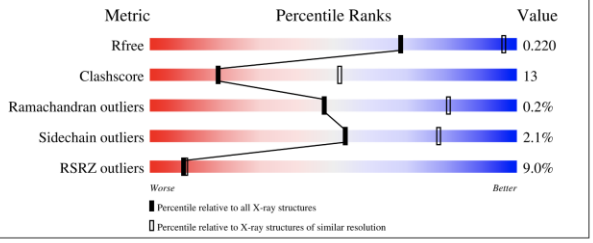
12. **Lines 462-463: Why have the crystallographic models been re-refined?**

    Following the initial quality control check, we noticed that the geometry and density interpretation in the original structures had a number of outliers, therefore we re-refined them to avoid carrying the distorted geometry over to our model. In case of 5JK7 we only re-refined the part of the model that we used for interpretation of our structure. We have now prepared an additional table summarizing the validation statistics after re-refinement (**Table EV2**). Below we show side-by-side comparisons of the newly refined model versus the original model using the PDB validation protocol.

    2HYE

Statistics after re-refinement:	Statistics Original model:
	

5JK7

Before	After:
	

13. **Some references seem incomplete (e.g. Hrecka 2007, Marks 2016, Rohou 2015, Stark 2010, Wang 2016, Yang 2020, Yu 2013, Zhang 2001)**

    We apologize for this software error. We have now corrected all references.

14. **Fig. 5B is not referenced in the text.**

    We apologize for the omission. Figure 5B is now exchanged to Figure 5A and is referenced on line 292.

15. **Fig. 5C and S5C are mis-labelled: UNG2 is actually Vpr and vice versa**

    We thank the reviewer for pointing out the mix-up. We adjusted Figure 5C and Figure EV 5D accordingly.

16. **Fig. S1A: please explain "sensitivity score" in the legend**

    We now include a description of the sensitivity score in the legend (line 931), as follows: “The dropout score in the shRNA screen was converted into log fold-change (logFC) per shRNA per cell line. The logFC was then normalized per sample to obtain a shRNA level sensitivity score. The shRNA level scores are further aggregated to gene level sensitivity scores using either the ATARiS algorithm (Shao et al., 2013) or the RSA algorithm

(König et al., 2007). These give a measure of the statistical significance of the drop-out of those 20 shRNAs used per gene compared to the remainder shRNAs in the screen (McDonald *et al*, 2017)"

17. Figs. S1B, D, S2A, B, C, are not referenced in the text.

We apologize for the omission. All figures are now properly referenced in the revised text as follows: Figure EV 1B on line 111, Figure EV 1D on line 132, Figure EV 2A, B, and C on line 142.

18. Fig. S4A left panel: control trace (N8-CRL4DCAF1 FL) is missing

In order to compare the enzyme activity towards different substrates, one needs to divide the Vmax by the enzyme concentration used in the reaction. For N8-CRL4^{DCAF1}-VPR-UNG2 we had to increase the CSN concentration to 150 nM in order to observe a signal. Yet, we only needed 15 nM of CSN to observe a signal for the wildtype N8-CRL4^{DCAF1} or N8-CRL4^{DCAF1}-MERLIN and using 150nM, on the other hand, gave a rate that was too fast to measure. In order to compare enzyme activity in this particular case, it is thus necessary to compare the K_{cat} values not the Vmax. As evident in the table below, while V_{max} values are more or less similar, but it would be misleading to compare the V_{max} values, as we used different enzyme concentrations. We now include the following sentence in the legend: "We used the K_{cat} value to compare the enzyme activity towards the different substrates due to different enzyme concentration being used (15 nM vs. 150 nM) and K_{cat}= Vmax/[enzyme]"

Substrate	**[CSN] (nM)**	**V_{max} (μM/s)**	**K_{cat} (S^{-1})**	**K_m (μM)**
N8-CRL4^{DCAF1}	15	0.009	0.6027	0.735
N8-CRL4^{DCAF1}-VPR-UNG2	150	0.006	0.04062	0.963
N8-CRL4^{DCAF1}-MERLIN	15	0.008	0.5372	0.661

.

Referee #3:

Mohamed et al. reported biochemical and structural characterization of Cul4-Rbx1-DDB1-DCAF1 complex and revealed an autoinhibited tetrameric configuration that is regulated by Nedd8 modification and substrate binding. SEC-MALS experiment and a low-resolution Cryo-EM model confirmed a tetrameric configuration of Cul4-Rbx1-DDB1-DCAF1 complex. Prior study showed that LisH motif in DCAF1 could form a dimer. With this knowledge, the authors docked the modelled DCAF1 armadillo and LisH structures into the density. The model shows that LisH domain forms a dimer and the dimer further assembles into a tetramer via intermolecular Rbx1 and DCAF1 WD40 domain interaction. A VRSA-loop in the WD40 domain contacts Rbx1 RING domain. Alanine substitution of the conserved R1247 in this loop disassembled the tetramer into dimer confirming the importance of this interaction. The tetrameric configuration is not compatible with Ub-loaded E2 or Nedd8-loaded E2 binding. The authors showed that WT-complex had reduced autoubiquitination activity and Ubc12-mediated Nedd8 modification of Cul4 compared to R1247A-complex suggesting that tetramer represents the inactive conformation. Prior studies showed that Nedd8 modification of Cullin induces conformational changes in the C-terminus of Cullin and Rbx1; the conformational change would abolish Rbx1 and WD40 interaction. Consistent with this notion the authors showed that Nedd8-modified complex exists as a dimer. Moreover, the tetrameric model is incompatible with CSN interaction and the authors showed that CSN did not pull down tetrameric form of the CRL4-DCAF1 complex. Lastly, the authors showed that substrate, Vpr-UNG2, binding to the WD40 domain of DCAF1 converts the complex into a dimer, whereas non-substrate binding such as MERLIN had no effect on the tetrameric assembly.

This study reveals a novel autoinhibited tetrameric configuration of CRL4-DCAF1 complex where Rbx1 is partially occluded from binding to E2-Ub or E2-Nedd8. The autoinhibition is released when CRL4 is Nedd8 modified or when DCAF1 engages a substrate. These findings will be of interest to EMBO Journal readers.

I have few comments that need to be addressed.

1. **The authors mentioned that CRL4-DCAF1 complex exists in a dimer/tetramer equilibrium, but SEC-MALS shows a single mono-disperse species. What is the protein concentration of CRL4-DCAF1 complex used in the SEC-MALS experiment? Is the tetramer formation concentration dependent**

   This is very good point. We loaded 2 mg/ml of protein on a 24-ml Superose 6 column in the SEC-MALS experiment. At the concentrations we screened for cryo-EM and negative-stain EM (0.05-0.2 mg/ml), we could still only observe a tetrameric complex in the micrographs. We concluded that the complex likely exists in a dimer-tetramer equilibrium due to the observation that it can be neddylated, yet with a slower rate, and that neddylation likely requires a dimer intermediate for access. Overall, the equilibrium of the unneddylated $CRL4^{DCAF1}$ complex appears to be largely shifted on the tetrameric state, with a tight Kd. It is correct that we did not detect the dimeric intermediate directly, and only infer its presence by the ubiquitination and neddylation data. Following the

suggestion of the reviewer (detailed below), we now also added the possibility that the ubiquitin/NEDD8-loaded E2 could more actively assist in the tetramer opening. We have now added the following sentence to the discussion section on line 353, and it now reads: “Moreover, although our findings are consistent with $CUL4^{DCAF1}$ being in a dimer-tetramer equilibrium, we cannot exclude that Ub/Nedd8-loaded E2 complexes, or in fact also CSN, possess additional molecular mechanisms to open the $CUL4^{DCAF1}$ tetrameric complex and thereby convert it to a dimer.”.

2. **In the Cryo-EM analysis, is the complex predominantly tetramer or is there a small fraction that is dimer?**

   We could observe many broken particles in $CRL4^{DCAF1}$ micrographs. However, we did not identify any classes that looked predominantly dimeric. The tetrameric form of the $CRL4^{DCAF1}$ ligase appears to be the dominant form.

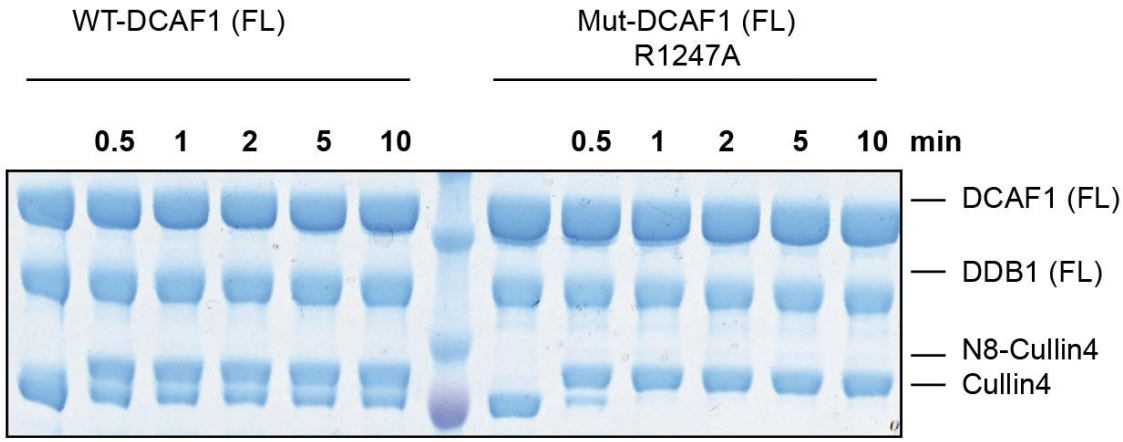
3. **The model showed that the tetrameric configuration is incompatible with E2 binding, but CRL4-DCAF1 is active in ubiquitination and neddylation suggesting that it can still bind E2. The authors suggest the CRL4-DCAF1 is in a dimer/tetramer equilibrium and it is the dimer population that is active. Could addition of Ub-loaded E2 or Nedd8-loaed E2 compete with WD40-RING interaction in the tetramer and cause the disassembly of tetramer? In a way, the tetrameric configuration reduces the binding affinity for Ub/Nedd8-loaded E2?**

   The data presented in the revised manuscript supports higher ubiquitination and neddylation rates with the tetramerization deficient mutant than with the largely tetrameric wild-type (**Figure 3C-D, and Figure 4B)**. Based on the structural models, we hypothesize that access for a Ub/Nedd8-loaded E2 is somewhat restricted. We did not obtain, and could not commercially source, sufficient amounts of Ub/Nedd8-loaded E2 to test their effect on $CRL4^{DCAF1}$ oligomerization. The ubiquitination results and the finding that unneddylated $CRL4^{DCAF1}$ is tetrameric under all conditions tested in EM and SEC-MALS, unless mutated, further strengthens the model that $CRL4^{DCAF1}$ is present in a dimer-tetramer equilibrium. Having said this, we cannot exclude that Ub/Nedd8-loaded E2 complexes have special, exchange-factor-like functionalities that could assist in opening the dimer. We thank the reviewer for making us aware of this possibility and have added the following sentence to the discussion on line 353 and it now reads as follows: “Moreover, although our findings are consistent with $CUL4^{DCAF1}$ being in a dimer-tetramer equilibrium, we cannot exclude that Ub/Nedd8-loaded E2 complexes, or in fact also CSN, possess additional molecular mechanisms to open the $CUL4^{DCAF1}$ tetrameric complex and thereby convert it to a dimer.”.

4. **In Figure 4B the loading of middle two panels are not the same. CRL4-DCAF1 (R1247A) bands seem less intense compared to WT panel. It is evident that there is a faint un-neddylated CRL4 R1247A band, which will show up if loaded similarly as WT panel. It would be useful to compare the initial rate of Nedd8-CRL4 formation using shorter reaction time point since at 5-min time point neddylation is nearly complete.**

We thank the reviewer for suggesting this experiment. A figure panel is now included that shows the comparison of the initial rates of N8-CRL4^{DCAF1} WT vs. mutant at shorter time points of 30 sec to 10 min (**Figure 4B, left panel**). In this experiment, we were actually able to capture the time when a fraction of the dimeric mutant is still unneddylated (at 30 seconds). So, while the CRL4^{DCAF1} dimeric mutant is fully neddylated after 1 min, the wild-type tetrameric CRL4^{DCAF1} is almost fully neddylated only after 180 min (**Figure 4B**).

Regarding the sample loading, we include here the upper part of the gel, where the loading of the WT vs. mutant samples can be judged taking into account the different components of the CRL4^{DCAF1} complex, such as DDB1 and DCAF1. From this, we are convinced the amount of protein loaded is largely the same for the WT and the mutant.



5. **In Figure 5A and 5D, CRL4-DCAF1-VPR, CRL4-DCAF1-VPR-UNG2 and CRL4-DCAF1-MERLIN elution profiles have a large peak near the void volume. Do these complexes form larger oligomer? Please explain.**

   The peaks near the void contain heterogenous and aggregated protein complexes, and in some cases may also contain traces of nucleic acids. We now include the MALS data that contain the Molar Masses of the protein fractions at/near the void volume in Figure EV 5A, B, and C. As evident from these traces, the void contains large aggregates with poor polydispersity. As such void behavior is often observed with protein complex purification, and non-stoichiometrically assembled CRLs, we did not further investigate this behavior. We now added a sentence to the methods identifying these as high molecular, polydisperse, aggregates. See line 452: “The protein fractions at/near the void volume elute as heterogenous high molecular weight aggregates, and we therefore did not further investigate their behavior (**Figure EV5 A, B, and C**).”

6. **In Figure S4 why is initial rate Vo expressed as uM/s? This is different from kcat (s-1). In Figure S4B, the y-axis numbers are mislabelled? kcat is ~0.6 s-1 but the curve is showing 0.006.**

   In order to compare the enzyme activity towards different substrates, one needs to divide the Vmax by the enzyme concentration used in the reaction. For N8-CRL4^{DCAF1}-VPR-

UNG2 we had to increase the CSN concentration to 150 nM in order to observe a signal. Yet, we only needed 15 nM of CSN to observe a signal for the wildtype N8-CRL4^{DCAF1} or N8-CRL4^{DCAF1}-MERLIN and using 150nM, on the other hand, gave a rate that was too fast to measure. In order to compare enzyme activity in this particular case, it is thus necessary to compare the K_{cat} values not the Vmax. As evident in the table below, while V_{max} values are more or less similar, but it would be misleading to compare the V_{max} values, as we used different enzyme concentrations. We now include the following sentence in the legend: "We used the K_{cat} value to compare the enzyme activity towards the different substrates due to different enzyme concentration being used (15 nM vs. 150 nM) and K_{cat}= Vmax/[enzyme]"

Substrate	**[CSN] (nM)**	**V_{max} (μM/s)**	**K_{cat} (S^{-1})**	**K_m (μM)**
N8-CRL4^{DCAF1}	15	0.009	0.6027	0.735
N8-CRL4^{DCAF1}-VPR-UNG2	150	0.006	0.04062	0.963
N8-CRL4^{DCAF1}-MERLIN	15	0.008	0.5372	0.661

Thank you for submitting your revised CRL4-DCAF1 study to our editorial office. I have now heard back from the three original referees, and given their positive re-reviews (copied below), we shall be happy to accept the study for EMBO Journal publication, following incorporation of the remaining minor referee comments and the following editorial points.

REFEREE REPORTS

Referee #1:

The authors have addressed all of my concerns. They have improved the manuscript substantially.

Referee #2:

Mohamed et al. contribute a highly relevant structural and biochemical study of $CRL4^{DCAF1}$, demonstrating a novel mechanism of CRL4 autoinhibition mediated by tetramerisation. I recommend the revised manuscript for publication, with only a few minor comments:

Fig. EV3A DCAF1 ARM is coloured brown while the legend states yellow.

Fig. EV3C, D could be swapped since EV3D is mentioned first in the text. Also, the confidence score of the alphafold model should be colour-coded in EV3C, maybe as additional panel.

Fig. EV5E The N8-$CRL4^{DCAF1}$ (FL) blue control trace is missing in the plot (left panel)

Referee #3:

The authors have addressed all my concerns and the manuscript is suitable for publication.

The authors have made all requested editorial changes.

Accepted **10th Sep 2021**

Thank you for submitting your final revised manuscript for our consideration. I am pleased to inform you that we have now accepted it for publication in The EMBO Journal.

EMBO PRESS

YOU MUST COMPLETE ALL CELLS WITH A PINK BACKGROUND ↓

PLEASE NOTE THAT THIS CHECKLIST WILL BE PUBLISHED ALONGSIDE YOUR PAPER

Corresponding Author Name: Nicolas Thomä
Journal Submitted to: The EMBO Journal
Manuscript Number: EMBOJ-2021-108008

Reporting Checklist For Life Sciences Articles (Rev. June 2017)

This checklist is used to ensure good reporting standards and to improve the reproducibility of published results. These guidelines are consistent with the Principles and Guidelines for Reporting Preclinical Research issued by the NIH in 2014. Please follow the journal's authorship guidelines in preparing your manuscript.

A- Figures

1. Data

The data shown in figures should satisfy the following conditions:

- ➔ the data were obtained and processed according to the field's best practice and are presented to reflect the results of the experiments in an accurate and unbiased manner.
- ➔ figure panels include only data points, measurements or observations that can be compared to each other in a scientifically meaningful way.
- ➔ graphs include clearly labeled error bars for independent experiments and sample sizes. Unless justified, error bars should not be shown for technical replicates.
- ➔ if n< 5, the individual data points from each experiment should be plotted and any statistical test employed should be iustified
- ➔ Source Data should be included to report the data underlying graphs. Please follow the guidelines set out in the author ship guidelines on Data Presentation.

2. Captions

Each figure caption should contain the following information, for each panel where they are relevant:

- ➔ a specification of the experimental system investigated (eg cell line, species name).
- ➔ the assay(s) and method(s) used to carry out the reported observations and measurements
- ➔ an explicit mention of the biological and chemical entity(ies) that are being measured.
- ➔ an explicit mention of the biological and chemical entity(ies) that are altered/varied/perturbed in a controlled manner.

- ➔ the exact sample size (n) for each experimental group/condition, given as a number, not a range;
- ➔ a description of the sample collection allowing the reader to understand whether the samples represent technical or biological replicates (including how many animals, litters, cultures, etc.).
- ➔ a statement of how many times the experiment shown was independently replicated in the laboratory.
- ➔ definitions of statistical methods and measures:
  - common tests, such as t-test (please specify whether paired vs. unpaired), simple $\chi 2$ tests, Wilcoxon and Mann-Whitney tests, can be unambiguously identified by name only, but more complex techniques should be described in the methods section;
  - are tests one-sided or two-sided?
  - are there adjustments for multiple comparisons?
  - exact statistical test results, e.g., P values = x but not P values < x;
  - definition of 'center values' as median or average;
  - definition of error bars as s.d. or s.e.m.

Any descriptions too long for the figure legend should be included in the methods section and/or with the source data.

In the pink boxes below, please ensure that the answers to the following questions are reported in the manuscript itself. Every question should be answered. If the question is not relevant to your research, please write NA (non applicable). We encourage you to include a specific subsection in the methods section for statistics, reagents, animal models and human subjects.

B- Statistics and general methods

Please fill out these boxes ↓ (Do not worry if you cannot see all your text once you press return)

1.a. How was the sample size chosen to ensure adequate power to detect a pre-specified effect size?	Not applicable.
1.b. For animal studies, include a statement about sample size estimate even if no statistical methods were used.	Not applicable.
2. Describe inclusion/exclusion criteria if samples or animals were excluded from the analysis. Were the criteria pre-established?	Not applicable.
3. Were any steps taken to minimize the effects of subjective bias when allocating animals/samples to treatment (e.g. randomization procedure)? If yes, please describe.	Not applicable.
For animal studies, include a statement about randomization even if no randomization was used.	Not applicable.
4.a. Were any steps taken to minimize the effects of subjective bias during group allocation or/and when assessing results (e.g. blinding of the investigator)? If yes please describe.	No.
4.b. For animal studies, include a statement about blinding even if no blinding was done	Not applicable.
5. For every figure, are statistical tests justified as appropriate?	Not applicable.
Do the data meet the assumptions of the tests (e.g., normal distribution)? Describe any methods used to assess it.	Not applicable.
Is there an estimate of variation within each group of data?	Not applicable.

USEFUL LINKS FOR COMPLETING THIS FORM

http://www.antibodypedia.com
http://1degreebio.org
http://www.equator-network.org/reporting-guidelines/improving-bioscience-research-repor

http://grants.nih.gov/grants/olaw/olaw.htm
http://www.mrc.ac.uk/Ourresearch/Ethicsresearchguidance/Useofanimals/index.htm
http://ClinicalTrials.gov
http://www.consort-statement.org
http://www.consort-statement.org/checklists/view/32-consort/66-title

http://www.equator-network.org/reporting-guidelines/reporting-recommendations-for-tumo

http://datadryad.org

http://figshare.com

http://www.ncbi.nlm.nih.gov/gap

http://www.ebi.ac.uk/ega

http://biomodels.net/

http://biomodels.net/miriam/
http://jjj.biochem.sun.ac.za
https://osp.od.nih.gov/biosafety-biosecurity-and-emerging-biotechnology/
http://www.selectagents.gov/

Is the variance similar between the groups that are being statistically compared?	Not applicable.

C- Reagents

6. To show that antibodies were profiled for use in the system under study (assay and species), provide a citation, catalog number and/or clone number, supplementary information or reference to an antibody validation profile. e.g., Antibodypedia (see link list at top right), 1DegreeBio (see link list at top right).	These citations/catalog numbers were provided in the Materials and methods section of the revised manuscript.
7. Identify the source of cell lines and report if they were recently authenticated (e.g., by STR profiling) and tested for mycoplasma contamination.	

* for all hyperlinks, please see the table at the top right of the document

D- Animal Models

8. Report species, strain, gender, age of animals and genetic modification status where applicable. Please detail housing and husbandry conditions and the source of animals.	Not applicable.
9. For experiments involving live vertebrates, include a statement of compliance with ethical regulations and identify the committee(s) approving the experiments.	Not applicable.
10. We recommend consulting the ARRIVE guidelines (see link list at top right) (PLoS Biol. 8(6), e1000412, 2010) to ensure that other relevant aspects of animal studies are adequately reported. See author guidelines, under 'Reporting Guidelines'. See also: NIH (see link list at top right) and MRC (see link list at top right) recommendations. Please confirm compliance.	Not applicable.

E- Human Subjects

11. Identify the committee(s) approving the study protocol.	Not applicable.
12. Include a statement confirming that informed consent was obtained from all subjects and that the experiments conformed to the principles set out in the WMA Declaration of Helsinki and the Department of Health and Human Services Belmont Report.	Not applicable.
13. For publication of patient photos, include a statement confirming that consent to publish was obtained.	Not applicable.
14. Report any restrictions on the availability (and/or on the use) of human data or samples.	Not applicable.
15. Report the clinical trial registration number (at ClinicalTrials.gov or equivalent), where applicable.	Not applicable.
16. For phase II and III randomized controlled trials, please refer to the CONSORT flow diagram (see link list at top right) and submit the CONSORT checklist (see link list at top right) with your submission. See author guidelines, under 'Reporting Guidelines'. Please confirm you have submitted this list.	Not applicable.
17. For tumor marker prognostic studies, we recommend that you follow the REMARK reporting guidelines (see link list at top right). See author guidelines, under 'Reporting Guidelines'. Please confirm you have followed these guidelines.	Not applicable.

F- Data Accessibility

18: Provide a "Data Availability" section at the end of the Materials & Methods, listing the accession codes for data generated in this study and deposited in a public database (e.g. RNA-Seq data: Gene Expression Omnibus GSE39462, Proteomics data: PRIDE PXD000208 etc.) Please refer to our author guidelines for 'Data Deposition'. Data deposition in a public repository is mandatory for: a. Protein, DNA and RNA sequences b. Macromolecular structures c. Crystallographic data for small molecules d. Functional genomics data e. Proteomics and molecular interactions	This section is now included in the revised Manuscript.
19. Deposition is strongly recommended for any datasets that are central and integral to the study; please consider the journal's data policy. If no structured public repository exists for a given data type, we encourage the provision of datasets in the manuscript as a Supplementary Document (see author guidelines under 'Expanded View' or in unstructured repositories such as Dryad (see link list at top right) or Figshare (see link list at top right).	
20. Access to human clinical and genomic datasets should be provided with as few restrictions as possible while respecting ethical obligations to the patients and relevant medical and legal issues. If practically possible and compatible with the individual consent agreement used in the study, such data should be deposited in one of the major public access-controlled repositories such as dbGAP (see link list at top right) or EGA (see link list at top right).	
21. Computational models that are central and integral to a study should be shared without restrictions and provided in a machine-readable form. The relevant accession numbers or links should be provided. When possible, standardized format (SBML, CellML) should be used instead of scripts (e.g. MATLAB). Authors are strongly encouraged to follow the MIRIAM guidelines (see link list at top right) and deposit their model in a public database such as Biomodels (see link list at top right) or JWS Online (see link list at top right). If computer source code is provided with the paper, it should be deposited in a public repository or included in supplementary information.	

G- Dual use research of concern

22. Could your study fall under dual use research restrictions? Please check biosecurity documents (see link list at top right) and list of select agents and toxins (APHIS/CDC) (see link list at top right). According to our biosecurity guidelines, provide a statement only if it could.	Not applicable