Dear Editors,

Please consider our revised manuscript of "Symplectin evolved from multiple duplications in bioluminescent squid".

Comments from Reviewer 1 (Sabrina Pankey)

Experimental design

- COMMENT: Details on BLAST parameters missing. Need to report e-value thresholds or minimum sequence similarities.
- REPLY: e-value of 10^{-10} was used, this has been added to the text.
- COMMENT: Some details relevant to Figure 2 are missing or unclear: this tree was generated by BLASTing genomes and transcriptomes? Was the same BLAST threshold used for Figures 1 and 2?
- REPLY: The trees are the same tree, where Figure 1 is a detailed version of Figure 2. This has been clarified in the captions of both figures.

Comments for the Author

- COMMENT: l.42 More background in intro: What roles do pantotheinase and biotinidases play in the well-characterized models?
- REPLY: We added the statement "Such proteins have been characterized in mammals, and in *Drosophila* and have roles in recycling of the enzymatic cofactors biotin and pantothenic acid."
- COMMENT: l 60. Methods: how were samples preserved (-80, RNAlater etc)?
- REPLY: All samples were flash-frozen in liquid nitrogen. This has been added to the text.
- COMMENT: l.103: Avoid speculation in results
- REPLY: The statement "suggestive of secondary loss" has been removed.
- COMMENT: Table and l. 112: Uroteuthis 'enopla' should be Uroteuthis edulis.
- REPLY: Changed, and in the text as well.
- COMMENT: l.97-113: Big caveat to testing hypotheses of molecular evolution with transcriptomes is that failure to detect a paralog's transcript does not imply its genomic absence. Must be careful to avoid this interpretation in stating results: For instance,some species are represented by multiple tissues..this could artifactualy increase the chance of detecting symplectin paralogs.
- REPLY: We were aware of this problem. The statement "likely due to coverage limits of transcriptome sequencing" has been changed to "likely subject to coverage limits of transcriptome sequencing or tissue-specific expression."
- COMMENT: l.127-148: Avoid speculation in results
- REPLY: This paragraph has been moved into the discussion.
- COMMENT: l.164 define 'dhCtz'
- REPLY: This was in the introduction, but was missed by both reviewers. We have clarified that this is an abbreviation.
- COMMENT: Figure 1. colors confusing: purple has two meanings: symplectin paralog and 'other proteins' mechanism?
- REPLY: Good consideration, we have changed the colors.
- COMMENT: Fig. 2: Is this meant to demonstrate that that symplectin is a member of the vanin family (and thus justify using the vanin structure for modelling), with pantotheinase as the outgroup?

- REPLY: Vanin-1 is the only member of this superfamily that has a crystal structure. This has been clarified in the text.
- COMMENT: There are no support values to indicate this hypothesis is supported.
- REPLY: This node, among many, was not supported, as stated in the caption. The relationships of groups within this family cannot be determined based on the data we have, as noted in the text on the subject of duplications.
- COMMENT: The color scheme is confusing. Color schemes mis-match between Fig 1 and 2. Either make symplectin group same color in both for clarity, or use entirely different palettes for each figure. Might be easier to follow authors' interpretation if rooted tree is rooted at choanoflagellates. In the legend, remind reader that btd=biotinidase and vanin=pantotheinase
- REPLY: The caption now notes that vanin is pantotheinase. Because branch support is low for backbone nodes, we felt it was better to show the tree as unrooted, otherwise readers are likely to question why arthropods are sister group to all other animals, among other problems.

Comments from Reviewer 2

Introduction

- COMMENT: Line 42. Do pantetheinases and biotinidases belong to the same superfamily, and if so, what is the name of this superfamily and its general characteristics? This is useful introductory information from a symplectin structural point of view.
- REPLY: The sentence in the introduction has been expanded to say "Previous work highlighted the sequence similarity of symplectin to members of the biotinidase/pantetheinase family, part of the superfamily of carbon-nitrogen hydrolases".

Materials and methods

- COMMENT: Line 80. Include the accession number for the symplectin sequence used.
- REPLY: Added.
- COMMENT: Lines 88-89. Include the PDB accession ID for the vanin-1 structure used in modelling here at first mention (even though it is mentioned elsewhere).
- REPLY: This has been changed.
- COMMENT: Line 94. What was the cut-off used to define homologs in the BLAST search?
- REPLY: This has been added to the text.

Results

- COMMENT: Line 116. In comparisons... please re-phrase this sentence to make its meaning clearer.
- REPLY: This has been changed.
- COMMENT: Line 117. State the % amino acid sequence identities of the known sequence of symplectin with these two proteins.
- REPLY: Added.
- COMMENT: Consider moving the first reference to Figure 2 from line 139 to earlier in the subsection Symplectin-like proteins across metazoans, to help readers better interpret the associated results.
- REPLY: Done.
- COMMENT: Lines 122-4 Because ... loss. This statement is possibly a little strong for the evidence available. Consider adding likely or probably to due to secondary loss.
- REPLY: We have changed this to "...absence of this protein family in hemichordates and ctenophores is likely a secondary loss."

- COMMENT: Line 178. Please comment on how trustworthy this model is, eg using the % identity between symplectin and vanin-1 and perhaps any measure of confidence the HHPred program provides.
- REPLY: Only one step provides an evaluation number, but all three model evaluation files can now be downloaded with the model at the bitbucket repository. We have noted this in the methods.
- COMMENT: Figure 4. The structure in Fig 4A is nicely pictured, but please label residues mentioned in the text as well as C390 in this image, including the catalytic triad of the nitrilase domain, to enable the reader to picture where these residues are positioned to each other. The residues in Figs 4B and C are hard to see. Please make them different colours to the backbone cartoon representation, and label them.
- REPLY: We have added labels to parts B and C.

Discussion

- COMMENT: Line 196. It may be more accurate to name this subsection Catalytic structure predictions, or something similar, since the function of the protein is already known (catalysis of bioluminescence), and the subsection speculates on which residues may be involved in catalytic activity.
- REPLY: This has been changed in the text.
- COMMENT: Lines 201-202 Although two other luciferases have solved crystal structures... This sentence is inaccurate and needs adjustment and elaboration. The authors possibly mean that the structures of two other luciferases that use coelenterazine have been solved. Actually, there have now been three different types of coelenterazine-utilizing luciferase structures solved, the Oplophorus and the Renilla luciferases referenced here, and also the hydrozoan photoproteins (aequorin, obelin and clytin). For more details on the Renilla and hydrozoan types of structures, see the review: Sharpe, Hastings, and Krause (2014) Luciferases and Light-emitting Accessory Proteins: Structural Biology. DOI: 0.1002/9780470015902.a0003064.pub2 A photoprotein is actually an intermediate-bound luciferase enzyme that is accumulated in the absence of a final reactant (calcium), and is then discharged rapidly when calcium is added (producing a flash of light), so should be included in luciferase discussions. ...these were unbound forms so mechanistic generalizations cannot be made Actually, it has been shown that aequorin uses the triad of tyrosine, tryptophan and histidine for catalysis, and the three main catalytic residues in Renilla luciferase are thought to be aspartic acid, histidine and glutamate. Speculation on the catalytic residues of symplectin should take this information into account.
- REPLY: We prefer to stand by our original statement that possible comparisons are limited at the moment. We had examined and cited several papers of the structures, and we disagree with the interpretation that E-H-D form the catalytic triad of the Renilla luciferase. Loening et al 2006 present a structure where the the residues closest to the decarboxylated carbon of coelenterazine are N309 and D162. Although Renilla luciferase does not hold coelenterazine as a peroxide like aequorin, we consider it unlikely that E-H-D are capable of catalyzing the reaction on the opposite end of the molecule. Thus, either E-H-D is not the catalytic site, or the coelenteramide-bound structure that was presented does not represent the product-bound state of the enzyme. Structures with bound coelenterazine were obtained for aequorin, but this is not the case for Renilla or Oplophorus luciferase, so no obvious commonalities could be identified.
- COMMENT: Line 203. Consider labelling these residues in Fig 4A.
- REPLY: Done.
- COMMENT: Line 228. Please elaborate on what derived means in this context.
- REPLY: In retrospect, this was a confusing sentence. We have changed this paragraph to say "First, one or many of the cephalopod proteins may still have biotinidase activity or act as a hydrolase in other contexts via the nitrilase domain. Since symplectin is therefore predicted to have two separate functional domains, it is possible that symplectin performs multiple functions, and may even still have a role in biotin metabolism. Secondly, if the ancestral function is a biotinidase, then the pantetheinase activity of vanin-1 and related proteins found only in vertebrates therefore is a derived function."
- COMMENT: Line 266. Please provide more detail on what these proteomic investigations would entail, so the reader can judge if they will be useful.

- REPLY: This sentence has been changed to “For this reason, a more directed approach of proteomic investigations of bioluminescent material from certain species may prove to be more useful, particularly for a species like the vampire squid, which has a secreted bioluminescence”

Comments from Reviewer 3

- COMMENT: Please consider adding the common name of the squid Sthenoteuthis oualaniensis (purpleback flying squid?) to help enable more potential readers to find the paper during literature searches.
- REPLY: This has been added into the introduction.
- COMMENT: Line 12. Do you mean four distinct groups of these proteins? Please clarify.
- REPLY: This has been changed to “four well-supported groups of proteins are found in cephalopods, one of which corresponds to symplectin.”
- COMMENT: Line 15. Insert catalysis after pantetheinase, and of these after all.
- REPLY: Done.
- COMMENT: Line 42. Check spelling of pantetheinase/pantotheinase throughout manuscript.
- REPLY: The enzyme is pantetheinase, and produces pantothenic acid. This has been fixed in the text.
- COMMENT: Line 69. Insert the before Qiagen.
- REPLY: Done.
- COMMENT: Line 189. Change addition to additional.
- REPLY: Done.
- COMMENT: Line 192. Add the before nitrilase.
- REPLY: Done.
- COMMENT: Line 211. Remove a from in front of another.
- REPLY: Done.

Thank you for your consideration,

Dr. Warren Francis, PhD
University of Southern Denmark
Campusvej 55, Odense M 5230, Denmark