**COVER LETTER**

**Submission ID**: 129

**Title**: Comparative Visualization of Protein Secondary Structures

**Date**: July 17th, 2016

We would like to thank the anonymous reviewers for their helpful comments. We attempted to address all remarks in the revised version of the paper. In this letter, we give detailed responses to all reviewer comments and describe the corresponding changes in our manuscript.

**Reviewer 3 - primary**

1. *The case studies need to be significantly strengthened and better described. It needs to be clarified how the participants were selected and what their background was, along with the specific questions and tasks they were asked to address and carry out. Statements about the merit of the current tool relative to other tools need to mention the specific tools that are being compared. In addition, the authors should provide further clarification about the biological/structural insights claimed in the case study, as specifically requested by Reviewer 1. The authors should also address the question of whether and how the angles of features shown in the 2D view are useful to experts.*

TODO

1. *The evaluation of the gap insertion algorithm used in the alignment needs to be better explained. This should include explanation of 1) what it means for the method to fail, 2) what test set was used, 3) the failure rate.*

TODO

1. *The method of 2D encoding used in the juxtaposed and superposed views needs to be better described along with a discussion about scalability of the approach.*

TODO

As for the scalability, we added the following text to the paper:

The scalability of our approach highly depends on the input data and the similarity between the scrutinized chains. Theoretically there is no limit for the number of displayed chains, the only problem can be the readability of the resulting appearance. If the differences in the constitution and spatial orientation are small the approach can be used for dozens of solutions. On the other hand, when comparing significantly different solutions, the visualization will suffer from the occlusion problems even for a very small number of chains.

1. *The motivation for the specific visualization tasks that are required for the comparative analysis of proteins envisioned in this work needs to be elaborated and these tasks should inform the discussion of related work, which is currently not well focused and does not provide a clear case for  novelty.*

TODO

1. *Lack of strong motivation for specific visualization tasks. The “requirements” for the design of the tool are introduced on page 4 with little motivation for the specific tasks. It would be better if these tasks were introduced earlier, before the Related Work section, and were more clearly justified.*

TODO

1. *Description of Related Work is too narrow and doesn’t adequately  
   address the issue of novelty of the current work. The lack of a clear  
   statement of the visualization tasks to be performed seems to lead here  
   to a Related Work section that isn’t well focused on the issue of how  
   the current work differs from previous approaches. It is difficult to  
   assess from the description provided whether the proposed approach is  
   truly novel, and if so, how specifically?*

TODO

1. *The description of the gap insertion method is validated and the limitations of the approach lacks important details. On p. 6 it is stated that “The correctness was tested on many protein structures and only in some cases our greedy approach inserts a few unnecessary gaps into the chains.” Later in the Conclusions: “Insertion algorithm which, due to its simplicity, can in some specific cases insert too much gaps.” I think it would be helpful to show at least one example of how the method can fail and to give more specific information about the size of the test set used to evaluate the performance and the rate of failure. Even if these aren’t rigorous tests, more specific information would be helpful.*

TODO

1. *Methods for testing visualization are not adequately explained. Only biology background is given, but there are no specifics about what tasks users were asked to perform and how it was determine whether the objectives were attained. Furthermore, no information is given about the number or background of the “domain experts.” Rather than “case studies,” which implies a substantial degree of rigor, I think it is more accurate to describe what has been presented as “examples of use.” Finally, although it is claimed that the tool enables users to perform visual comparisons “more quickly and intuitively than before,” no information is given about what other tools are being referenced here. How does the information given by the present tool compare with that given by the Aquaria tool referred to in Figure 1?*

TODO

1. *It is not clear whether the 3D view can be zoomed to highlight differences in the regions being compared. This was perhaps just not mentioned in the examples shown, but it would seem to be an important feature, especially for performing comparisons of more than two structures. For example, in the example shown in Figure 10, the different colors used to distinguish among the structures being shown would seem to make it very difficult to see any highlighting of particular regions of the proteins. This would seem to limit the capability of the comparative analysis.*

The 3D view can be zoomed and also the selected parts can be highlighted (as can be seen in Figure 7). These features are definitely very important for the comparison, however, as also the reviewer points out, when comparing more structures with different coloring the highlights have a limited benefit.

So we added the information about the zooming possibility as well as the reference to Figure 7.

Minor points:

1. *Figure 1, right panel, page 2*

*Should label these structures and give some indication about what features are represented.*

We labeled the structures with their PDB identifiers and added more explanation about their selection and appearance.

1. *“It informs the user about the length and overall alignment of the compared structures", page 4*

*I don’t see how the length information is conveyed - at least there are no numbers indicating position along the sequence.*

We tried to clarify this by correcting the sentence in the following way:

“This part also gives the user the information about the relative length and global alignment of the compared chains. In other words, it aims to show the mutual positioning of the aligned chains which can be of different lengths.”

1. *Highlight, page 5*

*This could be better explained. What is the meaning of the offset in the start positions of the elements?*

TODO

1. “*The orientation of the secondary structures in the remaining aligned chains is adjusted according to the difference between the position and rotation of the corresponding secondary structures in the reference chain (see Figure 7)", page 5*

*More detail about how the offset and rotation are determined would be helpful.*

TODO

1. *“Secondary structures", page 6*

*What algorithms are used to determine these? I guess this is a standard part of protein visualization software, but it would seem helpful to have a reference.*

We use the DSSP algorithm to determine the secondary structures. We added the notion about this as well as the reference to the algorithm to the beginning of the Methodology section.

1. *Figure 8, caption. “…until both proteins are not processed”,  page 6*

*Check the wording. I think “not” could be removed.*

Corrected.

1. *In “Algorithm for Processing Molecular Dynamics, page 6*

*It seems strange to insert gaps in aligning proteins that have the same sequence and it seems strange to align features that appear at different places in the sequence. I think it would help to at least explain the reasoning here.*

TODO

1. *lenghts -> lengths, page 6*

Corrected.

1. *End of Methodology Section, page 7*

*Logical place to put reference to supplementary video, which is otherwise not referred to.*

We added the reference.

1. *“...can merge when the protein structures change. This merge is  largely caused by movements...", page 7*

*I this refers to shifts in position in the structure because no dynamics  
   are involved, but the language is a bit confusing.*

We tried to clarify this and we also added more information about the importance of presence of channels in the P450 proteins.

1. *“…using the juxtaposed views illustrated in the paper", page 7*

*It would help to refer to a specific figure here for comparison.*

We added the reference to the corresponding Figure 2 in the paper by Cojocaru et al.

1. “*the the", page 7*

Corrected.

1. *"The representation uses combines the advantages …", page 7*

*Sentence has numerous typos.*

Corrected.

1. *“When comparing many proteins or many time steps, the visualization starts to be too complex", page 8*

*Possible to be more specific about how many structures it is practical to compare?*

We tried to clarify this by adding the following text:

“Another issue occurs when comparing many proteins or many time steps. Even when the compared chains are very similar, the superposed visualization will be too complex at some point (i.e., there will be so many overlapped chains that the visualization becomes unreadable).”

The number of structures which is practical to compare is highly dependent on the differences between the compared chains. Therefore, we added the discussion about this issue to the end of the Results and Discussion section (see reply to comment no. 3).

**Reviewer 1**

1. *The approach presented in this paper first aligns multiple sequences with a  
   greedy gap insertion algorithm and then visualizes the alignment with  
   secondary structure ribbons and "adjusted orientations". This allows one  
   to overlay multiple sequence representations with encoding regions of  
   differing secondary structure composition. While the resulting figures  
   look promising, I would have liked to see more information on how exactly  
   angles for each of the arrows and ribbons are computed.*

TODO

1. *Also, I would like to see some discussion on the scalability of the approach: In the introduction, as the case studies show a maximum of 5 aligned structures.*

The discussion about the scalability of our approach was added to the end of the Results and Discussion section (see reply to comment no. 3).

1. *The supplementary video is well done and is very useful to demonstrate interactivity and that the system was successfully implemented. I would love this to be available online with source code made available to the public.*

The supplementary video is added to the final submission as well so we hope it will appear as a supplementary material at the conference web page. As for the release of the source code, we are prepared to send it on demand. And of course, we plan to release the online version as soon as possible.

1. *The first study is based on proteins with channels that 'can merge when the protein structures change,... largely caused by movements of specific secondary structures'. Is this merge visible in Figure 10? Are there any insights about the underlying structural changes in the protein family that can be inferred from the Figure, apart from the fact that some are well aligned (red box) and others are not (blue box)? Are angles useful to experts? If so, how are they using the angular information?*

TODO

Minor issues:

1. *As far as I know, the first work mentioning ribbon representations is:*

*[1] J. S. Richardson, “The Anatomy and Taxonomy of Protein Structure,” in Advances in Protein Chemistry, vol. 34, J. T. E. and F. M. R. C.B. Anfinsen, Ed. Academic Press, 1981, pp. 167–339.*

TODO

**Reviewer 2**

Negatives:

1. *poor English*

We tried to do our best to improve the readability and the language.

1. *better case studies or the details of how rhe case studies were chosen,    and how the work was evaluated would help.*

TODO

1. *addressing soem of the issues central to the approach like the gap issue*

TODO

1. *it wasn;t clear in the paper if the algorithm creates too many gaps or gaps that are too big - one of the issues with the English.*

TODO

**Reviewer 4**

1. *I missed a clear description and discussion of the 2D encoding used in the juxtaposed and superimposed views as well as a discussion of the scalability of the approach.*

TODO

1. *The manuscript frequently mentions domain experts, but it is not clear if they are co-authors or a different group of experts. Related to that, it might be more appropriate to refer to the "Case Studies" as "Usage Scenarios" if they were conducted by the authors themselves rather than by end users.*

We clarified this in the text by adding the following information:

“Our proposed visualization was tested on several usage scenarios proposed by the domain experts in biochemistry, namely in protein engineering. The group of experts consisted of one professor (head of the research group), two post-doc researchers, and two PhD students. All of them are very active in designing modifications of protein structures.”

1. *In the Gap Insertion Algorithm section, the authors say that they "propose a greedy algorithm with produces a sufficiently correct solution" but they don't define what they mean by "sufficiently". Also, further down they argue that "The correctness was tested on many protein structures and only in some cases our greedy approach inserts a few unnecessary gaps into the chains.". I assume this is correct, but the authors need to provide more information about this evaluation: 1. How was is done? 2. How are unnecessary gaps defined? 3. What was the gold standard? 4. How many are "some" cases?*

TODO

1. *In the last paragraph of the Methodology section contains a grammatical error: "except" needs to be replaced with "in addition".*

Corrected.